FACULDADE DE PSICOLOGIA E DE CIÊNCIAS DA EDUCAÇÃO

UNIVERSIDADE DO PORTO

AVALIAÇÃO PSICOLÓGICA II

TEÓRICAS

COLECTÂNEA DE TEXTOS

António Abel Pires

2017

THE CLINICAL PSYCHOLOGY HANDBOOK

Second Edition

Edited by
MICHEL HERSEN
University of Pittsburgh School of Medicine
ALAN E. KAZDIN
Yale University
ALAN S. BELLACK
Medical College of Pennsylvania at Eastern Pennsylvania Psychiatric Institute

PERGAMON PRESS
Member of Maxwell Macmillan Pergamon Publishing Corporation
New York • Oxford • Beijing • Frankfurt
São Paulo • Sydney • Tokyo • Toronto

CHAPTER 1

A PERSONAL HISTORY OF CLINICAL PSYCHOLOGY

Brendan A. Maher

This chapter is intended to be a personal history of clinical psychology.[1] With this comes a dangerous temptation to write as though the personal experience of the author has been systematically representative of the main themes that would be discerned in a comprehensive nonpersonal history of the field. At its worst this approach can mislead the reader into thinking that an anecdotal memoir is an adequate history; at best, it can employ personal experience to illuminate some of the main themes of the history. It can serve to describe not only what happened, but to give the flavor of what it was like to be involved in these happenings.

Clinical psychology is a creature that lives uneasily in the twin domains of academic science and the social world of the professions (especially psychiatry) in which political and economic factors play a significant role. Science does not develop free from the influence of nonintellectual factors; the professions, being conducted in the larger society, are even more subject to social movements and pressures. No history of clinical psychology, not even a personal one such as this, can be coherent without frequent reference to events lying outside the narrow boundaries of psychology itself. It has fallen to the lot of clinical psychologists of my own generation to witness far-reaching changes in the conception of clinical psychology. These changes have elicited different reactions. For some they have represented progress on the professional front; for others they represent an abandonment of valued goals. Few who have taken part in the field during these changes remain indifferent to them. The kind of history that is written inevitably reflects the reaction of the writer of the history. This is true of this chapter. Hence, in what follows, I have tried to weave together personal experience, aspects of the formal history of clinical psychology, and descriptions of the social milieu in which they occurred; I have also tried to make my own judgments clear where they exceed a simple description of events.

A personal beginning: My first opportunity to study

[1]The reader interested in more systematic histories of clinical psychology and of psychopathology may wish to turn to other sources such as Hilgard (1987), Kendall & Norton-Ford (1982), Kirsch & Winter (1983), Maher & Maher (1979), Maher W. B. & Maher B. A. (1985), Maher B. A. & Maher W. B. (1985), Maher (1988), and Matarazzo (1985).

psychology came and went when an occupational therapist suggested that I might while away my time as a plastic surgery patient in a Royal Naval Hospital by studying something. The year was 1945, World War II was over in Europe, and wounds received while minesweeping had removed me from the scene long enough to give time for reflection. Correspondence courses were offered to military patients by a government and college consortium; each of us could choose two. I chose psychology first, with meteorology as the alternate. Psychology proved to be oversubscribed, so I was assigned the meteorology course instead. By the time the first lesson was due to arrive, the authorities had moved me to another hospital, thereby terminating my prospects of becoming acquainted with either psychological systems or storm systems for the time being. It was not until I entered the University of Manchester two years later, in 1947, that I began to find out what psychology was all about.

MANCHESTER, 1947 TO 1950

Here psychology was one of the options from the list of subjects that could be studied in the freshman year. With English, history, and logic, it constituted my schedule for the first term. By the end of that term I had refined my interests down to English and psychology. Deciding to apply to take the honors curriculum in one or the other, I went to the English department, only to find that it was closed for lunch. As the passage of the war years had given me a sense of urgency about the future, I went directly to the psychology department. It was not closed for lunch, and so in this seemingly accidental fashion I entered the field that was to be the basis of my professional life. It was not an accident that I ever had real reason to regret.

The department of psychology at Manchester was dominated in its curriculum by its head, Professor Thomas H. Pear. Pear had studied at Wurzburg and was much influenced by the conceptions of that school of thought. Although he had written a small monograph on shell shock many years before, he was chiefly a social psychologist. His social psychology, and therefore ours, was informed by the writings of Gordon Allport, Margaret Mead, Ruth Benedict, and the social anthropologists generally. Pear's most visible interest was in the relationship between voice and personality—the most concrete interpretation of this being the link between accent and social class. He had conducted a large-scale survey with the aid of the BBC, in which radio listeners were asked to estimate the social background of speakers, each of whom read an identical passage from a scientific text. Impressive percentages of the British listening public correctly identified a Church of England vicar, a police officer, and so forth. An attempt by Pear to replicate this in the United States in collaboration with Gordon Allport had failed; it appeared that the boundaries of social class there were less indelibly marked than in the United Kingdom.[2]

No one in the department taught psychopathology or courses with a serious clinical content. A short series of visiting lectures on abnormal psychology was given by a faculty member from the Medical School. We were introduced briefly to the Rorschach cards in a session of a course defined as experimental laboratory psychology (the rest of that course was devoted to finding the "hot" and "cold" spots on the back of each other's hands) and to introspecting aloud to see if we could trap an imageless thought from time to time. Pear, we were told, had an antipathy to animal psychology of any kind and would not permit laboratory animals in the department. Hence—except for Thorndike—the works of Hull, Tolman, and the American behaviorists were largely unknown to us, although an intrepid junior lecturer once showed a short American-made film of rats running a maze, the accompanying commentary being delivered by him *sotto voce,* as if apprehensive that even the photographic image of the prohibited rodent would provoke professorial wrath. Nonetheless, even the most cursory examination of the shelves of the university libraries, or of the readings lists distributed in lectures, indicated that the most significant work in psychology was occurring in the United States.

We were, however, thoroughly exposed to the British school of intelligence testing and to the psychometric tradition generally. We read Spearman, became familiar with the concept of "g," and gave each other the Raven's Matrices test. Continuing study in philosophy and physiology forming part of the honors requirement in psychology, we spent time making nerve-muscle preparations under the benevolent eye of Professor Walter Schlapp and watching the wonders of a primitive syllogism computer developed by Wolfe Mayes of the philosophy department.[3] Whatever the strengths and weaknesses of that kind of

[2]This theme, of course, had inspired Shaw's *Pygmalion* and its later musical version, *My Fair Lady*.
[3]Turing was at Manchester at that time, and a colleague of Mayes, but the significance of all of this was, I regret to say, lost on this naive undergraduate.

instruction, it left an indelible impression that psychology was a science with bonds to both biology and philosophy and that this was to be welcomed rather than seen as a source of intellectual conflict.

Any understanding of the experience of those times is incomplete without reference to the world outside the university. A Labour government had come to power in Britain with a large majority of seats in the House of Commons and with an agenda intended to put right many of the grosser social and economic injustices of the prewar years. A keystone of the new social edifice was the National Health plan, and close to that was the intention to reform those aspects of the educational system that prevented all but a few children from entering the college preparatory high school system. Times were hard in Britain. Although the war had been over for two years, rationing of many commodities was still in force, especially of heating fuel, clothing, and some foods. Rental housing was almost nonexistent, more than half a million homes having been destroyed by air raids and by other consequences of war. Manchester had been heavily bombed in the Christmas air raids of 1940, and new building was slow to take place.

The social purpose of the times was in keeping with the historical ambience of the university and of the city of Manchester itself. Engels's "classic slum," Salford, lay just across the city boundary. The black grime that coated every building in the city, including those of the university, was a silent reminder that this was one of the first centers of the Industrial Revolution in the Western world, and a reminder of the human horrors that had been created in the name of commercial success. Peterloo, the murder of protesting citizens by militia cavalry in St. Peter's Square in the city center in 1819, was permanently lodged in the communal memory of the city.

For many ordinary Britons, especially in the industrial North of England, there had been two war aims. The first was to defeat Hitler and the apparatus of fascism; the second was to bring into being a society in Britain that would distribute opportunity and power more evenly. The postwar electoral defeat of Winston Churchill, such a stunning surprise to the rest of the world, was less a surprise in a Britain in which this charismatic war leader had failed to understand that success in achieving the first war aim would not compensate for refusal to undertake the second.

The combination of the prospect of a brave new future of social justice, and the sharing of current hardships, combined to prolong the spirit of camaraderie of the war years. Psychology in this context offered a chance to contribute to the improvement of human life. As students of psychology, it seemed to most of us that we were engaged in a worthwhile human enterprise that would enable us to follow an interesting vocation with social meaning. One of the most applicable parts of psychology seemed to be clinical psychology—a helping profession.

Manchester University was also the first British university to establish a chair in American studies, the initial incumbent being a former Manchester graduate returning from Columbia University, Professor Kandel. Assisted by one junior lecturer, he gave courses ranging from the Federalist papers to the Federal Reserve Bank, from the geography of the Great Divide to the economics of the Great Depression, and from the politics of the Gilded Age to the music of the Jazz Age. Curiosity about America and the Americans had been stimulated in me by experience in joint operations with the U.S. Navy during the war, so I took the courses that were offered. Kandel was a wonderful teacher, sharpening one's interest in the culture of the United States although, as I discovered later, never fully managing to convey the effects of large distances and enormous variety that so influence American life.

During the final year of undergraduate study, it was necessary for us to face the issue of a postbaccalaureate career. Serious but brief further training in clinical psychology in Britain was available at the Institute of Psychiatry at the Maudsley Hospital in London and, to some extent, with J. C. Raven in Scotland. In both cases the training lasted little more than a year, emphasized personality assessment, and did not include the award of a higher degree. A more commonly available alternative was direct entry into the hospital system of the newly established National Health Service, where on-the-job training was provided on the apprenticeship model. Promotion through the higher grades would follow from length of service and demonstrated competence. The duties of clinical psychologists, however trained, were predominantly in the field of assessment and research. Psychotherapy by psychologists was essentially unknown, although this fact should be considered in the light of the parallel paucity of psychotherapeutic activities by psychiatrists in Britain.

Due to distinctions made between the practice of psychotherapy and the activity of child guidance, British psychologists employed within the educational system did practice techniques designed to change the behavior of children of school age, but this was not officially regarded as child psychotherapy—although it rather obviously met all of the definitional criteria for that activity.

All of this made it clear that if thorough graduate education in clinical psychology was to be found, it

would be necessary to go to the United States to find it. A fellowship from the Fulbright Commission and the English-Speaking Union combined to make it possible to do this. The letter bearing the news told me that I was bound for Ohio State University in Columbus; it is now a matter of retrospective embarrassment to me that I had to go to the atlas to find not only Columbus, but to find Ohio. Professor Kandel was not helpful in the matter, remarking only that the university there had a great football team. He was right about the football team. What he could not know was that circumstances were sending me to a center of intellectual ferment and debate in psychology that would have a critical influence on my own thinking for the rest of my professional career. I debarked from the Cunard liner *Mauretania* in August 1950, one of a group of such students on board, each of us struck by the incredible Manhattan skyline and many of us, unwittingly, a tiny part of the stream that was accompanying the shift of the intellectual center of influence of the Western world from Europe to the United States.

OHIO STATE UNIVERSITY, 1950 TO 1954

The department of psychology at Ohio State was large. Clinical psychology in 1950 consisted of a group of faculty including Julian Rotter, George Kelly, and Boyd McCandless. Paul Mussen and Alvin Scodel were to join the clinical faculty shortly thereafter. The rest of the department included, among others, Harold Burtt, the chairman, Paul Fitts, Donald Meyer, Harold Pepinsky, Herbert Toops, Robert Wherry, and Delos Wickens. Shortly before, Donald Campbell had departed for Northwestern University and Carl Rogers for Chicago. These were losses, but it was a department in its prime; I was lucky to be there.

The doctoral program in clinical psychology was designed explicitly to fit what was already being called the "Boulder model." In 1949, the year before my own arrival in Columbus, a major conference on the problems of doctoral training in clinical psychology had been held by the American Psychological Association (APA) at Boulder, Colorado. It recommended a model of doctoral training in which competence in the subject matter of general psychology and particularly in research design and statistics was an essential component of the training of clinical psychologists at the doctoral level, as well as training in clinical skills. Acceptance of this scientist-practitioner model was widespread in the leading universities in the country.

The Boulder conference itself was the outcome of earlier reports of APA committees in 1945 and 1947. These had proposed a four-year doctoral program to include one year of clinical internship to be taken either before or after the dissertation had been completed but before the degree could be awarded. A major impetus for these reports had been requests to the APA by the Veterans Administration (VA) and the National Institute of Mental Health (NIMH) for help in setting standards to be applied in their award of funds to doctoral programs designed to train clinical psychologists. By the early 1920s the shell shock of World War I had been recognized as a psychological response to stress. This had led to a greater emphasis on the provision of psychological services for veterans than had been the case in previous wars. World War II produced many more American casualties of all kinds than had World War I. This naturally created heavy demands on the services of the VA. Many new hospitals and clinics were being built in addition to the older structures established after World War I. Mental health professionals of all kinds were scarce, and it was necessary to train them if the needs of the patients were to be met.

In practice the application of the Boulder model at Ohio State meant that the first two years of doctoral training were spent in a mixture of courses and seminars in learning theory, physiological psychology, statistics, and so forth on the one hand, and courses in personality theory and clinical psychology plus practica in intelligence and personality testing on the other. The third year was generally spent as a clinical intern, and the fourth year was spent on the dissertation. Students being trained and supported by the VA generally alternated academic work and clinical assignments at a VA clinic or hospital in a sequence that provided the required total internship experience in a cumulative manner over the four-year span of graduate study. But much more important than the mechanics of scientist-practitioner training was the intellectual atmosphere in which it was conducted.

The fabric of clinical psychology graduate training presented a texture woven from several distinct strands. First and most important of these was a faith in the potential of behavioral science to put clinical practice on a firm base. The commitment to becoming a scientist-practitioner was not merely acquiescence in the long-standing academic requirement that the Doctor of Philosophy degree be a testimony to scientific and scholarly competence. It was due to a firm belief that then-current clinical techniques did not, in fact, possess adequate justification in scientific terms. This, it was felt, could be rectified by the systematic application of the scientific method and sophisticated

measurement procedures to the development of better techniques. Along with this went the observation that the training of psychiatrists in normal psychological science and in research methods was negligible.

If there were to be real improvements in clinical practice, they would have to come from a group of scientifically trained persons with firsthand experience of the clinical problems that were to be solved. These were to be the scientist-practitioner clinical psychologists. As the patients could not be expected to wait until the scientists had come up with effective procedures, the duty of the clinical psychologist was to apply what was available as skillfully as possible, but to do so with a conscious recognition that this was a temporary expedient, to be replaced with something much better as soon as possible. Creating the "something better" was a higher duty than mastering the techniques at hand.

One inevitable consequence of this was that entry into private practice was regarded as an undesirable career choice. Private practice appeared to border on the mildly unethical as it involved selling services of dubious adequacy and diverted the clinical psychologist from the main task of developing better methods. Indeed, too big a career investment in private practice might dispose the practitioner to reduce dissonance by defending the use of inadequate techniques rather than improving them.

A second thread on the loom of our experience was a sense of camaraderie. We were all engaged, we thought, in a great social enterprise. Scientific psychology would contribute to the postwar world; those of us involved in developing this psychology were joint contributors. By present-day standards, academic competition between us was minimal. Graduate students studied together, helped each other with shared notes, abstracts of required readings, and a host of other things designed to overcome the hurdles set by the curriculum; with a few exceptions we did not try to trip others so as to ensure that we got over the hurdles sooner or higher than our fellow students. Much of this was due, no doubt, to the fact that many of us were veterans and used to this way of life. But it was equally true of those who had not served in the military; the common theme was scientific optimism and the expectation of a useful collaborative future career in research, teaching, and clinical work. The emphasis was usefulness, not affluence. Recognition, if it came, was to be the result of genuine contributions, not of skilled public relations.

Third, there came an acquired indifference to the problem of diagnosis in psychopathology. Although we were exposed to the classic psychiatric diagnostic categories, it was implied that they were not really relevant to the understanding of abnormal behavior. From Julian Rotter we learned that maladaptive behavior could be understood in terms of social learning processes and that behavior change could be achieved by changing the client's expectancies. George Kelly was at work developing his theory of personal constructs and impressed on us the need to gain insight into the way in which clients perceived themselves and their social worlds. From both we learned that the task at hand was to develop a comprehensive theory of human personality and that the clinical implications would necessarily follow. Although our testing practica took us into the closed wards of the local state psychiatric hospital and the state school for the mentally handicapped, our experiences there were rarely linked systematically to the theoretical programs of the psychology department. Nor did the work of our mentors ever turn to consider the scientific problems of schizophrenia, the psychoses in general, or the classic neurotic syndromes.

Experimental studies of major psychopathology played little part in our graduate training at Ohio State, but there was opportunity and encouragement to undertake experimental work in the spectrum of psychology generally. It was because of this spirit that I was fortunate to obtain supervision and instruction from Donald Meyer as an assistant in experimental studies of the functions of the visual areas of the rat brain. Under the primitive conditions then prevailing, I was taught the techniques of cortical ablation and the conduct of experimental procedures for the investigation of discrimination learning in laboratory animals. Stereotaxic localization of the lesions was not then available. Errors in placement were determined by postmortem microscopic examination of the relevant brain sections. None of this had direct relevance to the clinical psychology or personality psychology that was then being provided in the doctoral program but was very much in keeping with my undergraduate education, in which physiology, psychology, and philosophy were assumed naturally to go hand in hand. Technical skills came first, and their theoretical integration would come later.

Psychoanalysis was dealt with critically at Ohio State. Its fatal shortcomings as a science were discussed, and we were in this way armed for the follies that we understood would await us when the clinical internship would bring us into daily contact with psychiatry—which in its academic medical settings was then effectively dominated by Freudian thinking and its variants. The arrival of Alvin Scodel did bring a more positive exposition of psychodynamic con-

cepts, but they remained generally on the edge of the main intellectual core of the clinical program.

In 1950 I met a fellow graduate student, Winifred Barbara Brown, then pursuing a demanding doctoral program in both clinical psychology and experimental psychology. From the outset she was, and has remained, my true companion and best friend. She provided for me an example of genuine and wide-ranging intellectual curiosity and personal and intellectual integrity. From Barbara I first learned the intricacies of Hullian learning theory as well as instruction in the habits and customs of American life, academic and otherwise. Our initial meeting at a department barn dance was followed by interactions in a testing course, in which the writer was required to simulate some syndrome of psychological handicap. From these and from evenings spent together studying and writing reports, our affections developed to the point at which we decided to marry in August 1952, at the beginning of the third year of graduate study, the time at which we were to enter internship. George Kelly, whose understanding of the constructs of others was not confined to theory, supported our applications for internship appointment at the Illinois Neuropsychiatric Institute in Chicago, where the department of psychology was directed by David Shakow. His support led to our acceptance together, and saw us, three days after the wedding, arriving in Chicago in a 1937 Chevrolet, hauling a rental trailer loaded with used furniture and books, accompanied by a Siamese cat and a mixture of enthusiasm and apprehension about the real world of clinical experience that lay just ahead.

David Shakow is undoubtedly one of the major figures in the history of clinical psychology in the United States. Born in 1901, Shakow had studied psychology at Harvard as an undergraduate. On graduation, financial need led him to obtain a position as a research assistant at the Worcester State Hospital in Massachusetts. Returning to Harvard as a graduate student in experimental psychology, he commenced work on a dissertation on subliminal perception under the supervision of E. G. Boring. The dissertation was unsuccessful, and Shakow returned to Worcester State Hospital in 1928. Here he began a major series of experimental investigations in the psychopathology of schizophrenia, many of them centering on the role of reaction time as an index of attentional deficit in schizophrenic patients. He did receive his doctoral degree from Harvard after the war, this time submitting a dissertation on schizophrenia.

Shakow's presence at Worcester State created a setting in which young psychologists could obtain internship experience during the prewar years.[4] His own professional mixture of quantitative experimentation and his personal psychoanalysis provided an early example of the values expected of the scientist-practitioner. In 1946, Shakow moved from Worcester State to Chicago, taking up a professorial appointment at the College of Medicine of the University of Illinois, of which the Illinois Neuropsychiatric Institute formed a component unit.

When we arrived in 1952, Shakow had assembled a talented group of colleagues in his department. Ann Magaret, Alan Rosenwald, and Charles Wenar provided training and supervision for five interns. Shakow's own work included the analysis of a large quantity of reaction time data from his earlier investigations at Worcester State Hospital.[5] However, he had also begun preparations for an ambitious project—the study of psychoanalytic therapy via film recording of actual therapeutic sessions.[6] This project required that a room be set aside for therapeutic sessions, equipped with lighting adequate to movie making. Although the recording of sessions was to be done with the knowledge and consent of both therapist and client, Shakow judged it advisable to render the camera unobtrusive. Hence, the room had been decorated with wallpaper, the design of which contained many whorls somewhat reminiscent of camera lenses, and one of which was. Nothing definitive appears to have come out of this study or subsequent studies at NIMH. The massive quantity of data was recalcitrant to analysis, and no conclusions were reached.

Ironically, in 1952, portents of future difficulties for the psychoanalytic hegemony of psychiatry were already present in several guises. One took the form of the traveling representatives of pharmaceutical com-

[4]One of these was Julian Rotter, who warned us that Shakow would expect not only dedication to the work of the internship but a willingness to take part in occasional performances of the operettas of Gilbert and Sullivan in his home. The warning proved correct on both counts.

[5]Shakow depended heavily in this work on the statistical skills of his research assistant, Jean Paulsen. We were delighted to be able to attend her wedding to Loren Chapman during that year.

[6]Hilgard (1987) described Shakow's studies as involving the videotaping of therapy sessions. This may have been the case at the NIMH, but his Chicago study employed a movie camera, video cameras and recorders being entirely absent from the department of psychology at that time.

panies. They were to be found in the lobbies and corridors of the hospital, distributing samples of the new drugs reserpine and chlorpromazine. Psychology interns wore long white coats while on duty, and being thus indistinguishable from residents in psychiatry received unsolicited samples of various drugs from detail men who did not feel obliged to ask us to show medical credentials. The psychopharmacological revolution had begun, and these were its first heralds.

Even as David Shakow was preparing his room for the study of psychoanalytic therapy, the *Journal of Consulting Psychology* published the now famous article by Eysenck (1952) reporting that patients treated psychoanalytically fared worse than those treated supportively by general practitioners, who in turn fared less well than those not treated at all. Eysenck's article was flawed by many problems related to the noncomparability of the groups treated and other such matters. Nonetheless, one conclusion was indubitable: namely that psychotherapy, especially psychoanalytic psychotherapy, had been applied for many years with neither any real evidence that it was effective nor any recognition that such evidence was essential to justify it. The paper provoked vigorous controversy within the ranks of psychotherapists generally, although the self-assurance of the psychoanalysts at the Neuropsychiatric Institute was such that its implications were largely ignored when it first appeared.

If developments in pharmacology and in the analysis of the outcome of psychotherapy stirred few ripples in the Neuropsychiatric Institute, the third portent of the early 1950s attracted even less attention. This was the rapid growth of knowledge of the neurobiology of consciousness, knowledge that would effectively demolish the belief that the study of conscious and unconscious mental life required an exclusive reliance on verbal behavior and other introspective techniques of the kind that formed the stock-in-trade of psychotherapy.

Each of these three beginnings has had lasting effects on the scientific study of psychopathology and hence on contemporary clinical work. In 1952, they appeared as clouds no bigger than a person's hand. Instead, we had the task of mastering the art of psychological assessment of both inpatients and outpatients. Projective testing, the Rorschach, and Thematic Apperception Test (TAT) being the primary instruments used, and intelligence testing with emphasis on the clinical interpretation of signs of pathology in responses to the Wechsler-Bellevue Scale, together with writing reports on the basis of these tests, formed the main part of our days. Regular seminars and journal meetings were interspersed with attendance at weekly staff meetings, these being conducted in the grand style.

Psychiatry at the Neuropsychiatric Institute was dominated by psychoanalysts, members for the most part of the Chicago Psychoanalytic Institute. The senior psychiatrists were all distinguished analysts; young psychiatrists in residency training were obliged de facto, if not de jure, to undergo a training analysis if they had hopes of a career in academic psychiatry later on. Franz Alexander, already famous for his psychoanalytic studies of psychosomatic disorders, was perhaps the leading light in this group of analysts, which included Roy Grinker, Beulah Bosselman, and Harold Solomon among others.

Wednesday staff meetings were attended by everyone. Residents, psychology interns, social work students, medical students, psychiatric nurses, and others filled an amphitheater looking down toward a central stage. Here sat the resident in charge of the case to be staffed, the psychology intern, and the social work student—in short the team responsible for assessing and treating the patient. In the front row sat the analytic professorate, the ultimate oracles in all professional and scientific matters. When the team members had finished their individual presentations, the audience listened to the opinions of the occupants of the front row. The consensus of opinion from that source determined the proper understanding of the case.

Experience of these meetings confirmed suspicions about the worth of the psychoanalytic enterprise conducted in this way. One staff meeting was spent discussing why a young female patient would be unlikely to leave the hospital in the near future due to some complicated intrapsychic factors. At the close of the meeting the resident remarked diffidently that the patient had already left the hospital to join her husband who had now found a job in another city; the meeting had been an exercise in futility, and the diagnosis was irrelevant to the factors that determined the patient's behavior. None of this disturbed the senior staff, who remained equally serene when asserting that sibling rivalry was the cause of disturbance in a patient who was an only child, and when providing a plausible psychodynamic account of a patient whose laboratory tests proved him to have brucellosis.

Here too, formal diagnosis was ignored except for statistical reporting purposes. It was necessary to assign some diagnostic label to each patient, and so the categories of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) were employed but had little role in the actual conceptualization of cases

by the clinicians who applied them. Perhaps the most consistent diagnostic decision related to a dichotomy of organic versus nonorganic. Tests for organicity were crude. Electroencephalography was available but was prone to many false negatives and false positives; X-ray methods were even less informative; and clinical neurological examinations were of limited validity. Under these conditions, support for the diagnosis of organicity was sought from psychological tests such as the Goldstein-Scheerer series and the subscale analysis of the Wechsler-Bellevue Scale. These formed prominent parts of the neuropsychological battery, and we spent much time in mastering the principles of their interpretation.

Rotation through the pediatric clinic gave experience in rapid clinical assessment of intelligence in children, three or four such assessments per day being a normal load. A month's assignment to the Cook County Psychopathic Hospital across the street gave first-hand experience of the overcrowding, understaffing, and generally inadequate services that were the lot of so many psychiatric patients—a stark contrast with the comfortable conditions prevailing in the Neuropsychiatric Institute.

There was also some time for research. Ralph Gerard was then director of the research laboratories at the Institute, symbolically placed in the basement that lay beneath the separate towers that housed psychiatry and neurology. With his encouragement and material support, it was possible to conduct a study of frontal lobe function in rats—a study stimulated in part by the then-current unfortunate wave of interest in psychosurgical intervention in the frontal lobes of psychiatric patients. A study of the Lowenfeld's Mosaic productions of arteriosclerotic patients in another institution, conducted collaboratively with a fellow intern, Anthony Martin, was also completed during the year.

Psychosurgery was not practiced at the Institute during our time there, although much interest in it had been engendered by the Columbia Greystones (Mettler, 1949, 1952) study of the outcome of psychosurgery. Electroconvulsive therapy (ECT) was in use and was applied to a wider range of patients than the depressive patients to whom it is now administered. Other than being assigned to observe the application of ECT, we had no direct involvement in the assessment decisions that led to its use in individual cases.

All in all, the internship year left lasting impressions. One was the reinforcement of the belief that technical advances were necessary and would not come without new knowledge. Psychopathology was the obvious discipline within which new knowledge would be organized, and hence the clinical psychologist must become a psychopathologist. A second impression was that as a contributing discipline to psychopathology, psychoanalysis was scientifically bankrupt, and that its clinical rituals had little relevance to the woes of the kind of patients who found their way into the wards of the Institute.

The question of psychotherapy provoked a kind of ambivalence that took some time to resolve. Psychotherapy at the Institute meant psychoanalytically based forms of treatment. Behavior therapy and nondirective forms of therapy were not employed. It was clear that the status differential between psychiatrists and psychologists in the Institute centered on who was judged competent to perform psychotherapy. Superior status was not based on a greater knowledge of scientific psychology, for the psychologists were well ahead of the psychiatrists on that score. The scientific contributions of David Shakow are with us still, but he was of lower status in the Institute hierarchy than were any of the psychoanalysts, whose contributions have long since been forgotten.

Nor was it due to the presumed ability of medically trained psychiatrists to provide medical services to psychiatric patients, for the medical problems of patients at the Institute were always handled by consultant physicians from the appropriate branch of medicine, not by psychiatrists. Some of the most embarrassing failures to realize that an apparently psychological problem had a major neuropathological basis were attributable to the hubris of psychiatrists; the clinical psychologists were much too aware of their own ignorance to take any risks. Medical training as such clearly had little relevance to the daily activities of psychiatrists.[7]

All of this being the case, why was psychotherapy forbidden territory to the psychologist? Much of the reason had to do with the policies of the psychoanalytic institutes in the United States, and much had to do with the fear that nonmedical psychotherapy would be the beginning of a movement that, by demedicalizing treatment, would lead to the ultimate exclusion of

[7]Given the psychoanalytic emphasis of the Institute, it was ironic to discover that Freud had suggested that medical training was unnecessary to the practice of psychoanalysis. What the psychiatric residents did have, which many psychologists did not, was extensive clinical experience of the mentally ill, but this was something that could be easily equalized without any concessions to the relevance of medical training.

psychiatry from the ancillary benefits of membership in the medical profession. Ambivalence about all this arose from a paradox. The paradox was that while admission to the practice of psychotherapy promised to diminish status differentials that had no justifiable basis in differential competence, this fact itself was a reflection of a more disturbing fact: namely, that there was no reason in evidence to support the notion that anyone should be doing psychotherapy. It was not at all clear how one person could be a more competent therapist than another unless there were skills to be learned that had a demonstrable relationship to successful outcome. No evidence of this existed. Someone had once defined psychotherapy as an indescribable technique applied to indefinable problems followed by unknown outcomes, adding the ironic comment, "For this rigorous training is required!" Thus, to lobby for the reduction of status differentials by demanding the right to do something irrelevant seemed silly and intellectually demeaning.

There were, of course, other solutions to the problem. Problem-specific treatments based on the systematic application of empirically validated psychological principles were being developed by the behaviorists, but these were still known more by rumor than by published report. Mention of them around the Institute elicited dismissive assurances that the only possible result of such approaches would be symptom substitution and that for that reason they were probably rather unethical.

However, there were already developments in the search for independent status as mental health professionals. In 1946, psychologists had become eligible for state certification in Virginia, and in 1951 a state licensing law had been enacted by the state of Georgia. None of this affected Illinois or Ohio at this time. For most clinical psychologists the most available and significant postdoctoral certification was voluntary and was provided by the diploma of the American Board of Examiners in Professional Psychology. Large-scale licensing was yet to come. Our own first identification with organized scientific-professional psychology came with our student affiliate membership of the APA in 1952. It gave us great pride and a consequent depth of disappointment in later years as the scientific purposes of the Association were so overwhelmed by the guild functions that it would gradually assume.

1953 saw Barbara and me homeward bound for Columbus and the final year of graduate study, confirmed in our beliefs about the irrelevance of psychoanalysis and in the attractions and challenge of the scientific mission of psychology. Our dissertations were directed at problems of the disruptive effects of anxiety on cognitive functioning, mine on the interaction of induced anxiety and personality factors on flexibility in problem solving and hers on the effects of trait anxiety on interference in verbal learning and recall. Graduation in June 1954 and the coincident expiration of my student visa sent me back to Britain immediately, where I was scheduled to be interviewed for a position as psychologist in an unspecified government department. Barbara, now expecting our first child, remained in Columbus to finish data collection, joining me in England in the fall of the year. Having reversed the traditional epic of the emigrant who leaves his home country penniless and returns to buy a Rolls Royce, I had returned to England with even less money in hand than the few pounds with which I had embarked on the *Mauretania* four years before. There was a pressing need for a job. I was able to satisfy the Civil Service interviewers that I was eligible for appointment to a psychologist position, but as days turned into weeks nothing was forthcoming about where the job would be and when it would begin.

BRITAIN, 1954 TO 1955

In the meantime I was offered a post as clinical psychologist at the Graylingwell Hospital in Chichester at a salary of £450 per annum, this being a 1-year increment over the starting salary for newly appointed BA apprentices, the increment being granted in recognition of the year's internship. It was accompanied by the gloomy prophecy of a local real estate agent that we might find housing in the area if someone should happen to emigrate, as there was nothing otherwise available. Martin Roth was the senior psychiatrist at Graylingwell at that time, and the post might have been attractive were it not for the fact that living would be difficult and that the kind of independent research activity that the doctoral program had prepared me for would not be available in the British hospital system. Hurried inquiry of the Civil Service brought the message that I had been appointed psychologist to Her Majesty's Prison in Birmingham, with an immediate assignment to the prison in Wakefield, Yorkshire, for a period of orientation to the work. A night train from Chichester to Manchester, and then on to Wakefield, and I was at last on the job and no longer a student. Due to the complexities of government regulations, I was regarded as officially working away from my home post of Birmingham, and hence received a significant additional salary allowance for the purely hypothetical disruption of residence. It came at the right time.

Wakefield Prison was a dark, castellated fortress, blackened by the same industrial grime that disfigured Manchester and most cities of the industrial North of England. Its visual grimness was somewhat alleviated by the daily sight of groups of prisoners, unescorted by guards, knocking at the main gate to be let inside after a day's work off the premises. The prison was atypical in that its inmate population included prisoners who had unsuccessfully pled the McNaghten defense to murder, but whose nonetheless evident psychiatric disability had led the Home Secretary to recommend to the Queen that their death sentences be commuted to life imprisonment. Not being legally insane, they could not be sent to any of the Broadmoor institutions designed for that purpose, but were imprisoned in a "normal" prison as the law required. Other prisoners at Wakefield had committed crimes that had some diagnostic implications, such as incest, child sexual abuse, and the "crime" of homosexual behavior—the law on that matter not yet having been repealed. Yet others seemed to have been sent to Wakefield as what the chief medical officer called the "Black sheep syndrome"—namely, that they came from the kind of upper socioeconomic background whose members rarely serve prison time for breaches of the law. One of these was the atom spy Klaus Fuchs, then nearing the end of his sentence. A silent man, he mixed little with other prisoners, was taciturn in our interview, and departed promptly for East Germany when released.

Little in my clinical training had prepared me for this work. An initial expectation that most of my clients would prove to be actually innocent, to have entered a criminal life as the victims of social injustice, or to have committed their crimes for reasons that lay deep in their psychopathology was quickly disappointed. True, some warranted the diagnosis of psychopathy, but of the others some were weak men driven to a single crime by unbearable frustration, and others were callously evil in the banal way described by Hannah Arendt.[8] Yet others were professional burglars and car thieves for whom occasional imprisonment was an occupational hazard, much as limb fracture is to a movie stunt person. All this was new to me. It was necessary to improvise on my clinical training if I were to develop any framework for understanding what was before my eyes. This, however, never came to pass. In March 1955, a few months after arriving in Wakefield, a cable from George Kelly inquired whether we could return to Ohio State immediately to replace a new young faculty member who had just been drafted.[9] And so, with our new baby daughter and our suitcases, we embarked on the aging Cunard liner *Samaria* at Southampton, rolled through a full gale off the coast of Nova Scotia, and arrived in New York just in time to begin the spring quarter at the university.

OHIO STATE, 1955 TO 1956

By 1955, the theoretical activities of both Julian Rotter and George Kelly had borne fruit in the form of major books. Rotter's *Social Learning and Clinical Psychology* was published in 1954 by Prentice-Hall. Kelly's two volumes, *The Psychology of Personal Constructs*, appeared in the following year, published by Norton. These works served as texts for the groups of students who worked with Rotter and Kelly respectively. Both books achieved for their authors greater visibility in clinical psychology as both, in different ways, offered systematic alternatives to the theoretical formulations then prevailing in clinical psychology. With the end of the academic year 1955 to 1956 my temporary appointment at Ohio State came to an end, and we left Columbus for the last time. I had received an appointment as Assistant Professor of Psychology at Northwestern University in Evanston, Illinois.

The years at Columbus had brought me into contact with a group of fellow students who were destined for later prominence: Rue Cromwell, Walter Mischel, and Lee Sechrest stand out, but there were many others both then and in the years that followed our departure.

NORTHWESTERN, 1956 TO 1958

Northwestern University in 1956 had a small but impressive department of psychology. Clinical psychology was staffed by Janet Taylor (later Janet Spence), Robert I. Watson, and myself, together with William Hunt, who was also department chairman. Other colleagues included Donald Campbell, Lester Clark, Carl Duncan, Donald Lewis, Benton Underwood, Leonard Diamond, and John Cotton. This was a collegial group, small enough to get together at coffee time each morning to discuss matters to do with psychology and science generally. Two topics came

[8]Hannah Arendt's well-known term for the psychological state of many of the people who operated the Nazi death camps.
[9]Jerome Kagan.

up repeatedly. One was the question of the epistemology of psychology and the related question of measurement. Campbell was the spark plug of these discussions, and our general interest in the matter led to lively colloquium presentations by visitors such Herman Feigl and Michael Scriven. MacCorquodale and Meehl (1948) had published their paper on the distinction between hypothetical constructs and intervening variables, and this distinction had provided an impetus for extensive examination of the philosophical bases of psychological science.

Our other discussions often centered on the importance of experimental psychopathology as the base of clinical psychology. Janet Taylor, then an assistant professor, was already a leading figure in that field, a visibility that had been initiated by her experimental work on anxiety and the development of the Taylor Manifest Anxiety Scale. Hull-Spence theory provided a framework for the study of psychopathological phenomena conceptualized as the product of drive (usually anxiety), the effects of reinforcement, generalization and extinction, and especially the interaction of the various kinds of inhibition (proactive, retroactive, and reactive). With this model in hand, the verbal behavior of schizophrenic patients was understandable as due to the interfering or facilitating effects of associations, the paired-associate learning task being the ideal vehicle for the investigation of this. Other schizophrenic anomalies lent themselves to interpretation as instances of an excessive range of stimulus generalization. Sarnoff Mednick had completed his doctoral dissertation at Northwestern on this topic.

Hullian influence was not confined to our studies of human behavior. Under the influence of the work of Neal Miller and his associates at Yale (e.g., Miller, 1948; Murray & Berkun, 1955), the employment of animal analogs to investigate the sequelae of conflict had become a significant element in our work. Beginning from the paradigmatic approach-avoidance conflict in rats, we looked at escape, displacement, the generalization of extinction (as therapy), and the like, together with the influence of early exposure to trauma (loud noise) on the adaptive capacity of the adult animal to produce these behaviors. The behaviors for which we sought animal analogues were drawn mainly from the catalog of psychoanalytic defense mechanisms and had evolved as a product of the famous seminar at Yale at which a synthesis of psychoanalysis and conditioning theory had been unsuccessfully attempted by Hull, Dollard, and others of their group. Although the immediate stimulus for the adoption of this strategy was the work at Yale, it is clear that the early work of Pavlov, and especially the production of experimental neurosis by Shenger-Krestovnikovna, were the real historical antecedents.

Implicit in the ready transition from human to animal and back again was the assumption that similar mechanisms were involved in all vertebrate species. This, in turn, meant that we had accepted, tacitly or overtly, the centrality of psychobiological processes in psychopathology. With this came additional questions, such as the effectiveness of psychosurgery and of the new tranquilizers in modifying conflict-induced behavior in animals. Looking back at that period with the benefit of hindsight, I am struck by the readiness with which we accepted a view of the unity of science. We did not worry much about the problems of biological reductionism, nor about any assumed conflict between social ("psychological") explanations and biological explanations of abnormal behavior. The task was to solve some definable human problems with whatever concepts and techniques proved to be relevant.

Perhaps the most important advantage of the intellectual community in the department at Northwestern was an atmosphere in which it was taken for granted that we were all psychologists first, and clinicians, social psychologists, or experimental psychologists second. Furthermore, as psychologists we were assumed to have not only a commitment to scientific psychology, but to have an interest in the other sciences generally.

LOUISIANA STATE, 1958 TO 1960

In 1958 we moved to Louisiana State University where, now an associate professor, I was to be responsible for the direction of the psychological clinic and the doctoral program in clinical psychology. The university was engaged in a major effort to recruit new faculty and to expand its programs, and this was particularly evident in psychology. Louisiana provided our first exposure to real segregation, to the many bizarre policies that had arisen to prolong it and the equally strange subterfuges that had to be developed to circumvent it. New Orleans was the main island of an archipelago placed in the surrounding sea of Southern conservatism; Louisiana State University was another, although there was a long road to be traveled before it could reach the present state of affairs. Consulting and supervisory duties required that I travel throughout the region, from Shreveport in the north to Gulfport, Mississippi, in the south, and reminded me again, as at Wakefield, how little I knew about the life circumstances of most people, and how confined our psychology was to understanding the

psychological workings of the minds of urban, middle-class, white adults.

In 1958, psychologists in Louisiana could be certified by a nonstatutory body. However, the state had recognized a need for trained psychologists when it inaugurated a program of fellowships for graduate students from anywhere in the country to study clinical psychology in any approved program in the country, provided that the fellow agreed to work in a mental health facility in Louisiana for a number of years after graduation. This program, together with the expansion of the faculty, brought a number of good students to the university. Several were to become leading figures in the direction of clinical psychology programs in the South, notably Henry Adams, Charles Noblin, and Roger McIntyre.

During this period several new developments were occurring at the national level. Behavior therapy was beginning to emerge as a force within clinical psychology. Wolpe's book, *Psychotherapy by Reciprocal Inhibition*, appeared in 1958, and Skinner had already described the possible uses of operant conditioning methods in the treatment of behavior deviations. My own research continued to explore the use of animal analogs. I saw nothing anomalous about my daily schedule, which typically included some hours in the animal laboratory followed by appointments with clients in the university clinic, together with teaching experimental psychopathology and psychological test interpretation. The hyphen in scientist-practitioner was a firm bond, and there was no good reason to try to pry it loose.

HARVARD, 1960 TO 1964

In the fall of 1960 we moved to Cambridge, Massachusetts where I had been appointed to a position as lecturer in the department of social relations. Robert White was the chairman, David McClelland directed the doctoral program in clinical psychology, and my immediate colleagues included Walter Mischel, Herbert Kelman, Irving Gottesman, Timothy Leary, and Richard Alpert. Our offices were located in a Victorian frame house at 5 Divinity Avenue, known as the Center for Research in Personality. It had once been the home of Walter Cannon, I was told. My own appointment was intended to provide a teaching replacement for Richard Solomon, who had recently moved to the University of Pennsylvania, leaving a gap in the provision of instruction in experimental and physiological psychology within the department. There were, of course, experimental and physiological psychologists in the department of psychology, but at that time the separation between the departments of social relations and of psychology was sharp, and hence Solomon had acted as resident experimental psychologist for the former. Thus, my teaching assignments included Theories of Learning, Proseminar in General Psychology (cotaught with Irving Gottesman), as well as a course that I had developed entitled Somatic Bases of Behavior and Its Pathology. Personality assessment was taught by Walter Mischel, and clinical practica were mainly in the hands of Justin Weiss at the Massachusetts Mental Health Center, or at the Veterans Hospital on Longwood Avenue, under the direction of Donald Ramsdell. Gottesman provided expertise in the Minnesota specialties of genetics studied via the twin method and the use of the Minnesota Multiphasic Personality Inventory (MMPI) and actuarial assessment procedures generally.

Clinical psychology at Harvard had been shaped by the influence of Henry Murray. Murray defined himself as a personologist. He was interested in the study of personality in normal individuals, his approach being informed by the concepts of psychoanalytic psychology much modified by his own theoretical framework of needs and presses. The approach to understanding an individual was essentially hermeneutic, having in it the portents of contemporary psychobiography and owing as much to the influence of the humanities as it did to behavioral science.

Social relations had been created for several reasons. Some of these had to do with the abrasive nature of relations between certain senior faculty members in the original departments from which it drew its founding members, but much of it grew from intellectual aspirations for a kind of unified study of human beings at all levels from the anthropological on down to the individual personality. In this respect, the work of Murray and his colleagues formed an important level in the edifice.

Unfortunately, there were serious problems inherent in the entire plan. One major problem lay in the fact that were no analogous departments of social relations at other universities. Students graduating with the doctorate from socal relations went on to develop academic careers in traditional departments of psychology, sociology, or anthropology elsewhere. In turn, there were no major journals of social relations, no grants for research given under a social relations heading, and no learned society devoted to it. Publication, grant applications, and society memberships were all firmly anchored under the traditional departmental rubrics. Hence, the realities of the academic and scientific worlds encouraged both student and faculty member to retain a disciplinary identity.

This meant that, by 1960, the department of social relations had developed into a de facto federation of Balkan states, held together rather loosely by certain department-wide requirements, but with no strongly held commitment to a central theoretical framework. For those of us in it, the arrangement had certain advantages. Interaction with anthropologists and sociologists on various committees, including thesis committees, created an opportunity for joint collaboration in research to an extent not often found within single departments. In my case it led to joint research into aspects of juvenile delinquency with a group of investigators coordinated by Stanton Wheeler, including James Q. Wilson, then at MIT, Walter Miller, Irving Zola of Brandeis University, and many others (Wheeler, 1968). Another collaboration involved work with the newly developed General Inquirer computer program. With this program, Philip Stone and others provided the means to categorize and analyze large quantities of verbal material; this meant that the detailed study of language utterances by psychotic patients was now possible and marked for me the beginning of a research enterprise that has continued ever since.

Others who were in the department in the period beginning in 1947, its early years (e.g., Korchin, 1983), describe it as more interdisciplinary than it appeared 13 years later. Gordon Allport, one of the founders of the department, continued to express throughout his career both the hope and the expectation that the social relations idea would be taken up elsewhere. By 1960 it was clear that this was not going to happen. In 1972, the department dissolved into its constituent parts and the experiment was over.

For clinical psychology, the organization of the department presented special problems, most of which arose from the fact that the personological approach with its emphasis on normal individuals had diminished the attention given to real-world clinical problems. No senior faculty member in the program had had significant clinical experience in a psychiatric setting. The clinical training of graduate students was left to adjunct faculty members in the associated hospitals and clinics of Harvard Medical School. There were clinically experienced senior faculty members such as the psychiatrist John Spiegel in the department, but they played no part in clinical training.

In 1962, Robert White concluded his term of office as chairman of the department, being replaced by David McClelland. Since McClelland was no longer available to direct the clinical psychology training program, I was asked to undertake the responsibility and was appointed chairman of the Center for Research in Personality. When I came into that office, the field of clinical psychology was on the verge of major changes.

THE 1960s, ANTIPSYCHIATRY, AND PROFESSIONALISM IN CLINICAL PSYCHOLOGY

This period was one in which the rise of what came to be called the antipsychiatry movement began. Erving Goffman published *Asylums* in 1961. Thomas Szasz published *The Myth of Mental Illness* in the same year, and other works were to follow in short order. Contained in this movement was a central message: Mental illness does not exist, in the sense that those who receive psychiatric diagnoses are not biologically ill in the way that people who receive a diagnosis of cancer or tuberculosis are ill. Terms such as *illness, hospital,* and *therapy* arise from the medicalization of what are essentially deviations from social, ethical, or legal norms of behavior. These problems would, and should, be treated differently were they not defined medically. However, the initial deviations are exacerbated by hospitalization in large total institutions which produce behavior that is a reaction to the hospital milieu. Reactions thus produced seem bizarre when viewed against the norm of the behavior of nonhospitalized persons, and so the initial diagnosis is confirmed, although it is the treatment that has led to the new "symptoms" that now appear.

From these basic propositions, several conclusions followed. One was that the use of coercive or biological therapies for mental illness (restraints, ECT, psychosurgery) were fundamentally misconceived. They might even be thought of as constituting punishment for deviant behavior, albeit disguised as treatment. Psychotherapy and psychopharmacology were seen as merely another form of restraint, subtler by virtue of the fact that they place the restraints inside the patient rather than outside. Another conclusion was that psychiatric patients, particularly involuntary inpatients, were victims of deprivation of civil rights, whose plight was in need of a political solution on a par with those then being sought for ethnic minorities. Ethical more than scientific matters lay at the root of these criticisms.

Yet a further implication of all this was that psychiatry itself, no matter how well intentioned its practitioners, functioned as a repressive force. To the obvious question "What is being repressed, and why is

it being repressed?" an answer was offered in its most explicit form by Ronald Laing (e.g., Laing, 1960). His answer was that psychosis, and especially schizophrenia, could best be understood as an inevitable consequence of the hypocrisies of family life in a capitalist society. Schizophrenic patients, he asserted, perceive the fundamental dishonesty of the social situation and are unable to enter into the self-deceptions that are necessary to adapt to it. Schizophrenics are therefore best understood as imprisoned dissenters from the prevailing dishonest ethos of their society. Just as the Soviet Union was using psychiatry to stifle dissent, so, according to the antipsychiatrists, did the Western world. Patients should be discharged from hospitals and readmitted, if at all, only voluntarily and briefly. Abuses abounded and were well documented by the antipsychiatrists. Reform was needed, but its nature was unclear, for the critics were at their best in pointing to problems, but less forthcoming about the solutions.

There were, of course, some answers to these complaints. Szasz's assertion that the major mental illnesses had no biological basis was premature and tantamount to the confirmation of the null hypothesis. None of the critics had spent much time looking into the history of the fate of the mentally ill before the establishment of the contemporary mental hospital to discover what their civil rights had been then. One little-noted response to the antipsychiatry movement was that it provided a convenient pretext for spending less money on the care of the mentally ill. From this perspective it could be seen as a refusal of the more affluent to contribute to the welfare of the poor.

At first, none of this had any discernible impact on the practical conduct of clinical psychology in the 1960s. However, it did reflect and contribute to a climate of opinion that was growing with the deepening involvement of the country in the undeclared Vietnam war. Established authority, many people now believed, was willing to override individual rights and to subvert democratic processes in the pursuit of its own goals. Appeals to patriotism, to communal duties, to obedience to the law, indeed to any conventional standard were seen as devices intended to keep people in order while those in power pursued their own immoral aims. Indeed, some concluded that if moral purposes were to be achieved by the individual, it would be by illegal actions such as the burning of draft records, the obstruction of research that had possible military applications, and the disruption of speeches by those who supported the establishment.

Some clinical psychologists concluded that the proper approach to mental health lay in preventive work in the social settings in which maladjustment developed. This view gave rise to the concept of community psychology. In 1963 Congress had enacted a bill (the Community Mental Health Centers Act) to establish a national network of community mental health centers. Two thousand of them were planned to be in place by 1975. This was in part an attempt to move the locus of treatment out of the hospitals and into neighborhood centers, but the implication was that the rationale for and types of treatment would remain unchanged. While this promised large-scale opportunities for the employment of psychologists, the continuance of traditional psychiatric approaches to mental health was a source of concern—hence, the movement for involvement in the community directly, via community psychology (e.g., Bennett et al., 1967).

For the most part, clinical psychologists in the 1960s were preoccupied with the need to secure professional independence from psychiatry. The act that had established the mental health centers had arisen from Congressional reports that also foresaw the introduction of a national health insurance system. This, in turn, meant that a crucial question would arise—namely, who would be entitled to direct reimbursement for the delivery of psychological services, psychologists in independent practice or psychologists working under medical supervision? Autonomous practice would require a legal definition of who was to be qualified to practice, and this meant that the motive to establish state licensing laws had become urgent.

As the opportunities for employment in clinical settings increased, there was an accompanying debate about the proper training for the clinical psychologist of the future. The debate turned on the necessity of the continuance of the scientist component of the scientist-practitioner model. Proponents of the abandonment of the model argued that research played little or no part in the responsibilities of psychologists in clinical settings and that the time spent in studying it could be better spent in getting more clinical experience. Those who recognized the necessity of research training in the curriculum for a Doctor of Philosophy degree proposed that a new degree be created, the Doctor of Psychology. Implicit in this latter suggestion was that there now existed a body of clinical techniques that were of proven effectiveness and that the mastery of them warranted the granting of a doctoral degree of some kind. In the discussions on the need for a separate degree, no significant data were presented relevant to this matter.

THE NEW THERAPIES: BEHAVIOR THERAPY AND HUMANISTIC THERAPY

The complex of influences that converged during the 1960s included not only antipsychiatry, antiestablishment attitudes, and pressures for professional autonomy. Equally prominent were the rise of the behavior therapies and the rise of the so-called humanistic therapies. Behavior modification had been applied to the training of the mentally retarded in the 1950s and was already in active use under the direction of Norman Ellis at the State Colony in Alexandria, Louisiana, during my own time at the University in Baton Rouge. However, the first event of note was the appearance of Wolpe's (1958) account of systematic desensitization, *Psychotherapy by Reciprocal Inhibition*. Although Dollard and Miller (1950) had attempted to formulate a learning theory translation of psychoanalytic concepts, it was no more than a translation and provided no significant new technique for the therapist. By the early 1960s, the operant conditioning approach had been applied to the modification of symptomatic behavior in chronic schizophrenic patients, with notable success.

At Harvard at that time there were no behavior therapists, and I was not to meet any until I moved to Wisconsin in 1964. However, James Holland of the department of psychology and I did give a joint graduate seminar in learning theory to students in that Department[10] and in this seminar Holland introduced us to the work of Allyon and his colleagues (Allyon & Houghton, 1962; Allyon & Michael, 1959).

Nor was there any specific exposure to the emerging range of new therapies, often referred to as humanistic. Serious discussion of the philosophical and methodological underpinnings of psychotherapy had been preempted by the unusual activities of my colleagues Timothy Leary and Richard Alpert. Space does not permit the full description of this period that it deserves; suffice it to say here that the notion of instant chemical access to insight and to ecstasy as opposed to the harder, longer grind required to master the profession and science of clinical psychology appealed to some students and repelled others. Leary and Alpert left the department of social relations at the end of the spring term 1963, Leary of his own accord[11] and Alpert by dismissal.

In my own case, the growing divide between the science and profession prompted me to start work on a textbook on psychopathology (Maher, 1966). From the beginning of my graduate studies I had been puzzled by the fact that textbooks on abnormal psychology, written by psychologists for psychology students, made no significant reference to psychological science. They were, by and large, watered-down expositions of Freudian psychoanalysis with a few case histories thrown in and a mandatory chapter on organic psychoses. On the other hand, the scientific journals devoted to abnormal psychology were amply supplied with articles describing experimental psychological and other quantitative investigations of patient samples. The Boulder model of training had not yet penetrated into the books that were to present the scientific material most central to our work.

WISCONSIN AND BRANDEIS, 1964 TO 1972

In August 1964 we moved to Madison, Wisconsin, where I had been appointed a professor of psychology. The department of psychology was strong in experimental and biological psychology; the roster of my colleagues included Wulf Brogden, David Grant, Harry Harlow, Fred Mote, and Michael Posner in experimental psychology, Leonard Berkowitz and Vernon Allen in social psychology, and Mavis Hetherington, Peter Lang, Barclay Martin, Albert Marston, and myself in clinical psychology. Carl Rogers had left Wisconsisn for the West Coast shortly before our arrival. It appeared that he had been unhappy with his move from Chicago to Wisconsin, feeling strongly that his students interested in clinical psychology, by

[10]Sometime after my arrival in Cambridge, Harvard had decided that the separation of psychology into two departments had adverse consequences for the curriculum, so I was one of four members of social relations then additionally appointed to the department of psychology as part of a plan to rectify this. Ironically my contribution to the department of psychology was to teach learning theories—a subject neglected in a department in which the operant paradigm dominated all alternative conceptions of the relationship of stimuli to responses. Thus in the process of teaching neo-Hullian ideas to the psychologists in Memorial Hall, I was first exposed to the literature of behavior modification, by an experimental psychologist. No such exposure occurred in the department of social relations.

[11]Most accounts of this episode state that Leary was fired by Harvard for his activities. In fact he simply left the campus, and his obligations, before the end of the spring term, 1963. He was granted leave without pay for the rest of the term, at which point his 5-year appointment as lecturer expired as scheduled.

which he meant the humanistic psychology of his own kind, were being faced with difficult and irrelevant requirements created by the experimental psychological culture of the Wisconsin department. Before departing he had communicated these concerns in a memorandum addressed to his colleagues, including myself still at Harvard.

There were concrete requirements in statistics and experimental design for all students, including clinical students. There was also a functioning psychological clinic in the department in which each clinical faculty member took responsibililty for the supervision of students. It was, in my judgment, a good example of the continuance of the Boulder model and intellectually congenial. However, Rogers's memorandum was another straw blowing in the wind, portending changes in clinical training at the national level. One year after our arrival in Madison, a conference was held in Chicago, the Conference on the Professional Preparation of Clinical Psychologists. I was invited to submit a position paper for the conferences and did so. It read, in part, as follows:

1. The most important functions of the clinical psychologist are in the areas known as behavior pathology, behavior deviation, etc. Derivative areas such as management counseling and motivational training are important in their own right; they are not central to the role of the clinical psychologist.
2. The basic purposes of the clinical psychologist are the prevention and cure of mental illness in both serious and minor forms.
3. Current psychological techniques of diagnosis and therapy are inadequate on two counts. First, the wealth of research evidence now available has cast grave doubts on their efficacy and validity; second, their application is too costly in manpower to support the belief that they may provide satisfactory solutions to the social evils of behavior pathology.
4. This technical inadequacy has generated two serious problems. One of these is the current state of low morale and doubt among clinical psychologists themselves. The other is the mounting pressure on the part of nonclinical psychologists to place clinical training in some other department or school, and to award some degree other than the Doctor of Philosophy on graduation.
5. Parallel problems exist in psychiatry where too early acceptance of psychoanalytic theory and therapy has produced a serious impasse . . . we must (therefore) attempt to expose the student to certain kinds of intellectual attitude. He should see his role as a potential innovator—not as a practitioner of existing techniques. He should be encouraged to develop an interest in problems, not methods . . . we should not be producing experts in the Rorschach or any other single technique. We need Jonas Salks, not Dr. Kildares. (Maher, 1965, p. 53)

The outcome of the conference was an endorsement in principle of the Boulder model, but with an important modification. While the clinical psychology student should be exposed to a series of topics, including research, it was necessary only that the student master a subset of them. In effect, the research requirement was now reduced to simple exposure, not mastery.

At Wisconsin the clinical faculty were attracted to the possibilities of the behavior therapies. They were deduced from a clear theoretical base, were consistent with what we knew of general psychological principles, and seemed to have a much better cure rate than the competing psychotherapies. Peter Lang was a central figure in this work. Our joint interest led us to take part in forming Section III of Division 12, the clinical division of the APA, the purposes of the section being to support presentations in the areas of behavior therapy and experimental psychopathology at the annual meetings of the Association. At this point we remained convinced that the steady drift of clinical psychology from its base in behavioral science need not continue and that a return to first principles could be achieved within the framework of the Association.

Others, more prescient, were less sanguine. The foundation of Psychonomic Society had already taken place a few years earlier, a response by an increasing group of research psychologists to what was seen as the gradual erosion of scientific values from the APA and an accompanying shift toward the protection of the guild interests of the clinical psychologist.

Several of us also took part in a 2-day workshop on behavior therapy given for the Wisconsin Psychological Association. By coincidence, the hotel at which the workshop was given was also the site of the annual meeting of the Wisconsin Psychiatric Association. It was immediately agreed that members of either society could attend either set of meetings. Consequently, our workshop audience now included a substantial number of psychiatrists, mostly in private practice. As a presenting speaker at the workshop, I was struck by the visible enthusiasm of the psychiatrists for the possibilities inherent in behavior therapy, in contrast to the defensive hostility shown by some of the psychologist-psychotherapists in private practice. It was a reminder that individuals whose daily living comes from the application of a specific technique are unlikely to examine it critically and equally unlikely to be hospitable to competing alternatives.

Discussions within the department of psychology frequently turned to the question of the possible appointment of a new faculty member in clinical psychology with full-time responsibility for clinical supervision and clinical teaching, but whose salary increments and promotions would not require evidence of a solid research contribution. The idea was that such an individual would replace the expected investment of time in research with an additional investment of time in clinical supervision. A broad spectrum of the department members was supportive of this idea; the allegedly anticlinical experimental psychologists were sympathetic to the need to be flexible in managing a clinical facility within an academic department. I gather that such an appointment was tried out, but this was after I had left Wisconsin for Brandeis, and I was not on hand to observe the success of the plan. To me, however, it was plain that Rogers had underestimated the support that the Wisconsin department would give to a scientific clinical psychology and perhaps overestimated the support that his own brand of humanistic psychology could command.

In the fall of 1966, with the textbook now in print, we traveled to Copenhagen where I was to be a Special Fellow of the National Institutes of Health (NIH) appointed to the Institute of Clinical Psychology at the university there. Professor Lise Østergård was the senior faculty member in clinical psychology and had a long-standing interest in language and thought disorder in schizophrenia. Several American psychologists were active in Copenhagen at that time. Sarnoff Mednick and Fini Schulsinger were deeply involved in the early stages of their longitudinal study of children at high risk for schizophrenia. Melvin Zax was in Copenhagen on sabbatical leave from Rochester, as was Oscar Parsons from Oklahoma, my office mate at the Institute. It was a congenial and productive year; it established a lasting affection for Danes and things Danish and a basis for two subsequent visits in later years.

In addition to collecting language samples from Danish subjects, I completed a short book on research methods in psychopathology (Maher, 1970) and an article to an early issue of *Psychology Today*, then newly published. During my time in the department of psychology in Memorial Hall at Harvard, I had made the acquaintance of George Reynolds, then a junior faculty member. George, Nicholas Charney, and another colleague were at that point formulating the concept that led to the founding of *Psychology Today;* it was to be for the educated layperson interested in psychology what the *American Scientist* is to the same kind of audience interested in the sciences generally. In its early years it managed to live up to that standard, although a major investment in elegant and creative color illustrations hinted at the possibility that concessions to popularity might someday outweigh faithfulness to science. Nonetheless, I was pleased with the article when it appeared until the discovery that the final rather flamboyant paragraphs had been added by a house editor.

While in Copenhagen I received an invitation to return to Massachusetts to take up a professorship at Brandeis. George Kelly had moved to Brandeis and was eager to establish a clinical program there in which the scientist-practitioner model would be maintained. Not long after accepting a formal offer from the university, I received from the department chairman Ricardo Morant a letter with the melancholy news that George had died. And so in the fall of 1967 we returned to New England to settle down in our present home after many years of academic and geographic moves, but without the much admired colleague whose presence in Brandeis had made the move attractive.

Brandeis was and is a small university. The department of psychology was therefore small in size and, of necessity, spread its faculty appointments across the range of specialties necessary to offer a satisfactory curriculum. At this time the dominant figure in the department was Abraham Maslow, then at the height of his career. His emphasis on the values of personal growth, and especially his concept of self-actualization, had wide appeal. By 1967 his reputation had placed him in the position of de facto leader of the humanistic psychology in the United States. Although the department did not have an APA approved program in clinical psychology, many of our graduate students had clinical interests. Courses and seminars, together with placements in clinical field settings, produced a small but steady stream of graduate students who went on to later positions of significance in clinical psychology and psychopathology. Douglas Whitman of Wayne State and Cathy Widom of Indiana provide obvious examples; the list could be expanded considerably with the names of Brandeis undergraduates who entered graduate clinical training programs elsewhere.

Other graduate students and some faculty were committed to the humanistic approach to psychology. The emphasis on the expression of personal feelings unconstrained by notions of tact, the rejection of personal privacy as a value in social life, and (as it has

appeared to me) the essential narcissism of the personal growth movement made an uneasy fit with the requirements of scholarship and research in a PhD program. Something of the spirit of the time has remained encapsulated for me in an incident in which a graduate student came to object vehemently to the grade of D that I had given to his exceptionally shoddy and ill-written term paper. "I deserve an A," he said angrily, "I had an A quality experience writing that paper." "No doubt," I responded, "but I had a D experience reading it, and that is what counts." But my satisfaction at the reply was irrelevant in an atmosphere where standards of competence or mastery were seen as oppressive and where the primacy of personal experience was taken for granted.

In some areas of clinical psychology the spirit was evident in the tacit decision that the demand for objective evidence of the efficacy of psychotherapy could safely be ignored. The experiences in therapy sessions were what therapy was all about—not some later outcome that was supposed to arise as a result of the therapy. The challenges of the statisticians and the behavior therapists could be safely ignored so long as the clients kept coming. Clients, in turn, could move from therapy to therapy until they found one that was congenial. The number of therapies now on the national scene increased steadily, and new labels proliferated. Thus, the primal scream, rolfing, and Gestalt therapy took their places in a kind of psychological supermarket, with issues of efficacy simply ignored.

The antipsychiatry movement continued to develop through the end of the decade. One controversial contribution to it was a paper by Rosenhan (1973) reporting a study in which it was said that a number of normal volunteer subjects managed to get themselves admitted to psychiatric hospitals by presenting minimal symptoms of auditory hallucinations, followed by perfectly normal behavior on the ward. The report stated that only the other patients recognized the normal subjects as normal, the staff failing to do so. After this report was published a flurry of debate in the professional journals followed quickly. Unfortunately, most of the questions raised could only be answered by inspection of the actual data of the study, and these have not been made available by the author in the subsequent 20 years. That the paper was so often cited in later years is perhaps a tribute to the strength of feeling about psychiatric diagnosis rather than the impact of convincing data.

However, events on the academic scene nationally were beginning to preempt concern. The second half of the decade was marked by student protests, mostly about the worsening Vietnam war, but also about the policies and practices of universities. So it was at Brandeis, and in the turmoil following the occupation of a building by minority students, that I was asked in 1969 to become dean of the graduate school. This was followed by appointment as dean of the faculty from 1971 to 1972. My personal involvement in clinical psychology diminished as a result of this, although I was able to maintain a continuing research effort in psychopathology. In 1972 I was invited to return to Harvard as professor, and thus entered my present position.

HARVARD, 1972 TO 1989

In 1972 the department of social relations had dissolved, and the department of psychology and social relations formed from the psychologists from the two separate departments. Sociologists had reestablished the department of sociology, and the social anthropologists had returned to join other anthropologists in their own department. The first chairman of the unified department of psychology was Gardner Lindzey, who had moved from the University of Texas at Austin to take over. Circumstances led him to leave Harvard as the first academic year was coming to a close, before the administrative task of making the paper unification of the department into an actual unity could properly be begun. Acting on the assumption that anyone who has been a dean should be able to administer a department, the president called on me to replace Lindzey as chairman, which I did in the spring of 1973.

By 1972, the formal doctoral program in clinical psychology was in the final stages of a phaseout. Not being present as a witness to the scene, my account of this is based on the reports of others. During the period from 1964 onward, the program had been permitted to run down in terms of faculty strength. In light of this, the university had agreed to appoint two new senior faculty members in clinical psychology, one to have a main emphasis on research and the other (to be appointed when the first was in place) to have a more clinical practical emphasis. Norman Garmezy was offered the first position and accepted it with the proviso that the program would be strengthened by one or more junior appointments. The department of social relations agreed to this condition—a necessary one if the program was to be restored to even minimal levels of strength. Apparently the first of the intended junior appointments was filled by a nonclinician at the instigation of the then chairman of the department, Garmezy not being consulted in the matter. Garmezy then resigned, naturally, and elected to stay at Minnesota. At this point the senior members of the department decided that the whole area of clinical psychology had become too problematic to be continued and began to dismantle it.

My own responsibilities as chairman were to create a unified department. They occupied much of my time. Coincident with my return to Harvard, I had accepted the editorship of the *Journal of Consulting and Clinical Psychology (JCCP)* for a 5-year period beginning in 1973.[12] This task, together with the chairmanship, meant that any serious effort at reestablishing clinical psychology in the new department would have to wait. There were enough potentially divisive matters within the department faculty without introducing an additional one.

In 1975 there began a steady decline in federal support for graduate training in psychology. As most doctoral programs were heavily dependent on this kind of support, the loss of stipends and associated expenses meant an inevitable reduction in the number of students. Coupled with this decline was a gradual tightening in the academic job market. In part a consequence of the completion of the expansion of universities that had taken place during the postwar period, it was exacerbated by an age distribution in the faculty skewed toward youth—with a consequent diminution of the percentage of retirements in the 1960s and onward. This has been the state of affairs up to the present time, although there are clear signs that matters will shortly improve.

While the number of entering graduate students in nonclinical psychology declined steadily, the total number of doctorates granted did not—the difference made up by the continuing rapid increase in production of PhD clinical psychologists. This increase antedated the 1970s, but the sharp increase in the ratio of clinical to nonclinical degrees did not begin until the mid-1970s. At Harvard, these events were reflected in a reduction of the entering class from 30 or more in 1972, to 12 to 14 at the present time. An inevitable consequence of the shortage of academic positions was that more and more doctoral graduates sought employment outside the academic and research fields, and that those already in nontenure academic positions were under greater pressure to obtain tenure.

This latter factor led to a steady increase in the submission of manuscripts to the *JCCP*. Between the years 1969 and 1972, submissions had declined slightly, from 708 to 624, a reduction of 11%. By 1977, total submissions had increased to 938, an increase of fractionally over 50%. Rejection rates stayed essentially steady at 80%, the overwhelmingly common reason for rejection being major methodological errors in the research reported. From the Editor's Annual Report for 1976, I quote the following paragraph:

> A reading of many manuscripts and all of the reviews strongly suggests that recent entrants into the field of clinical psychology are poorly educated in research design and statistics. Gaffes of a kind that were relatively rare in submitted articles in the late 1960s are now common and, more disturbing, correspondence with aggrieved authors suggests that many contributors are unaware of many of the most common criteria for validity, reliability, sampling, controls, etc. (p. 4)

In response to this problem we published a special methodological edition of the journal in 1978, which appears to have been widely used as a text. However, the proliferation of PsyD programs in free-standing professional schools and the reduced requirements for research training in many PhD programs constitute a tide against which no adequate dam could be erected, and there are no current signs of improvement in the general level of scientific sophistication in newly trained clinical psychologists in all but a few graduate programs.

By 1977, the department began serious consideration of the reestablishment of clinical psychology. What we had in mind was the scientist-practitioner model. As we looked into the requirements for APA accreditation, it became clear to us that the kinds of accommodation in our curriculum that would have to be made to meet these requirements could seriously compromise the intellectual adequacy of our doctoral standards. Matters had changed in the field of clinical psychology. Already more than half of the newly graduated clinical psychologists in the country were entering private practice as a main career actvity. The APA itself no longer appeared to serve our scientific interests, and many of us were discussing the need for reorganization of the Association if we were to remain in it.

For me the decision to withdraw from the Association was finally determined by its purchase of *Psy-*

[12]Due to the untimely death of my editorial predecessor, Jules Holzberg, before the end of his term, I had accepted responsibility for the remainder of that period, beginning the editorial office in 1972.

chology Today.[13] Catering to a perceived public need for sensationalized psychology, its ownership by the APA demeaned the image of a learned and professional society. At the time of purchase, the Association expressed the intention of improving the scientific quality and integrity of the articles. Unfortunately, this intention was futile when pitted against the hard fact that the necessity to publish between 80 and 100 such articles a year inevitably means that quality cannot be maintained. Psychology does not produce annually a sufficient number of new findings of general public interest to support this requirement. Instead, there was a continuing financial need to publish whatever was available—a concession to commercial aspects of psychology that was also evident in other areas of clinical psychology.[14] I withdrew sadly. My membership in 1952 had been a matter of some pride, and my disappointment correspondingly deep.

MARKET ASPECTS OF CLINICAL PSYCHOLOGY

With the increased emphasis on private practice, clinical psychology in the late 1970s moved toward a conscious attempt to market mental health. Some of this took the form of a growing literature in the psychological journals on the ways in which psychological services might be promoted. By the 1980s new journals had been formed devoted exclusively to the market interests of private practitioners. As a holder of a Massachusetts license, I found myself on the mailing lists of various derivative commercial enterprises devoted to the marketing of psychological services, or of services to psychologists designed to improve their marketing. Into my mail came letters such as the following: "Dear fellow professional: If your practice isn't where you would like it to be, now you can learn of a unique and exciting opportunity. My very first effort grossed $4800 working 1½ hours weekly for 16 weeks. That's $200 per hour." Yet another brochure offered me the opportunity to buy a franchise to provide an unspecified treatment for cocaine abusers, the cost of the franchise being based on the known average income level of the territory to be assigned. In case I did not understand the commercial advantages of franchising, the brochure included a simple account of the operation of McDonald's fast food service.

While the marketing of psychological services provided one way in which the practitioner could hope to increase income, the main source of payment for therapeutic services has been the payments provided by health insurance carriers.

Health Insurance

Most histories of clinical psychology attach importance to the resolution of the issue of the entitlement of psychologists to receive independent reimbursement for the provision of psychological services covered by health insurance systems. The major event in this history was the success of the Virginia Psychological Association in its court action to require the Blue Cross–Blue Shield organization to include clinical psychologists in its reimbursement policies when they are acting as independent health providers. The public argument of clinical psychologists was not that it opened the door to a significant source of income, but that it gave the client freedom to choose the mental health professional he or she preferred, and that the existence of licensing laws for psychologists meant that quality was assured.

Politically, this suit arrayed clinical psychologists against psychiatrists, who had been the only mental health professionals to receive insurance reimbursement as independent practitioners. However, the argument that by excluding psychologists from independent practice psychiatrists were acting in ways that limited competition was a double-edged sword. For example, in 1982, a new Florida licensing law was about to be enacted to replace a previous law that had expired in 1979. The APA, the Florida Psychological Association, and others lobbied extensively for passage of the new law. They were then sued by other mental health professionals (including social workers and other counselors) for having lobbied for a law that would reduce competition. The suit failed, but the implications were clear enough—namely, that the

[13]I was particularly disaffected by this action. Some years earlier, during my editorship of *JCCP*, the Association had distributed to a large sample of the membership a first issue of a proposed competing magazine to be produced by the APA. I received a copy. With the copy came a request for a vote on whether or not the Association should go ahead with the proposed magazine. The membership did not support the idea, and so it was dropped. Hence the decision to purchase *Psychology Today* seems to me to have been a clear rejection of what was known of membership sentiment on such matters.

[14]Public taste must have been underestimated, as the enterprise appears to have lost significant sums of money for the Association.

argument that competition must be opened up may be extended indefinitely to cover many other professional groups. Exclusion of these groups can only be justified if it can be shown empirically that their services are less effective, or their professional behavior less ethical than is the case with psychologists and psychiatrists. This has yet to be shown reliably.

At the time of writing, the policy of reimbursement of psychologists in independent practice has been extended widely. A more controversial issue, and one in which I was to become involved directly in a minor way, is the issue of mandatory coverage of outpatient psychotherapy in health insurance coverage. While the details vary somewhat from place to place, the mandatory provision in Massachusetts law then required all health insurance carriers to include at least $500 of coverage annually for outpatient psychotherapy. This provision, of course, does not specify who shall provide the psychotherapy, and hence psychiatrists and psychologists alike stand to benefit from the increased pool of dollars now set aside for their services. Before passage of the relevant act, coverage of outpatient psychotherapy was an option, chosen by the policyholder and, like all insurance options, subject to an increased annual premium.

In response to an article by McGuire (1981) I wrote a critical analysis of the whole concept of mandatory insurance coverage of psychotherapy (Maher, 1982). The core of the criticism is that unlike medical illnesses, the need for psychotherapy is determined mainly by the client; there is no independent laboratory test to determine the reality of the client's need (as there is in the case of tuberculosis, for example) and no assessor to decide that an insurable loss has actually occurred, as there is in the case of fire insurance. In brief, the fundamental requirement of insurance that the determination of need not be left in the hands of the policyholder is absent in the case of psychotherapy. While strongly in favor of the existence of voluntary optional add-ons of coverage for psychotherapy, I was, and am, opposed to mandatory coverage; it has seemed ironic that as clinical psychologists our insistence that insured persons should be free to choose between psychiatrists and psychologists has not been extended as logic requires to include the choice not to have coverage for psychotherapy at all.

There is no need to go into the details of the argument here; clinical psychologists of good will may have genuinely differing opinions on the merits of the matter. What was of more interest was the response that the article provoked. One letter writer (Reed, 1982) voiced his objections to the article. His letter ended as follows: "Such counterproductive, ruminative and misleading material is not appropriate for a journal representing an organization of professional and politically active clinical psychologists. Another such article and my council votes and those of a number of my associates will be applied to another division" (p. 17). Apart from the unusual notion that dissenting voices should not be heard in a professional journal, the writer had expressed his own right to discontinue his support for something that he did not wish to support—but that this right was to be denied to those consumers who did not wish to support the costs of psychotherapy.

It was clear that clinical psychology had evolved in a direction of which Albee (1970) had warned:

> The truly crippling sources of cognitive dissonance in the professional psychologist . . . are the fundamental differences between the scientist and the professional. . . . Science is open, and its knowledge is public. . . . Incisive mutual and self-criticism, replication, debate and argument over procedures . . . is ever present. In sharp contrast a profession must jealously guard its secrets. . . . Secrecy and mystery are essential. (pp. 1074–1075)

That the professional interests had now come to outweigh the scientific aspects of clinical psychology is evidenced by certain simple statistics, some of which are as follows:

1. In 1968 there was one professional PsyD program (at the University of Illinois); in 1982 there were 24 (McNett, 1982).
2. In 1968, 41.8% of all new PhDs awarded in psychology were in clinical psychology; in 1982 the percentage had risen to 52.6%.
3. In 1972 the percentage of all doctorate holders in the clinical behavioral sciences[15] who described themselves as self-employed was 11.9; in 1981 the comparable percentage had more than doubled to 26, while the absolute numbers had risen from 1,251 to 6,196 (Institute of Medicine, 1983).

[15]These figures are taken from the 1983 *Report of the Institute of Medicine, Personnel Needs and Training for Biomedical and Behavioral Research*. Clinical behavioral science in this report includes school psychology, counseling, and guidance as well as clinical psychology. Clinical psychologists form the large bulk of these numbers; the numbers do not include PsyD graduates, and hence the true numbers of private practitioners in 1981 are likely to have been much higher.

4. The research activities of clinical psychologists were surveyed in 1974 by Garfield and Kurtz (1974) and in 1981 by Norcross and Prochaska (1982). During the 8-year period between these surveys, certain trends were evident. The increasing opportunities for women in clinical psychology are indicated by an increase in the percentage from 16 to 21. The percentage of time reportedly spent in doing psychotherapy rose from 31 to 35, while the percentage of time reportedly spent in research remained low at a little over 7. Of this time, only 38% was spent on actually conducting research—the rest being described as "supervising research" or "scholarly writing." Thus, the percentage of working time actually spent in the conduct of research by clinical psychologists was 3, or roughly an hour and a half per the 48 hours that most reported as their normal working week. Clinical psychology as a field has continued to attract increasing proportions of women graduate students. In 1988 the number of doctorates awarded in all fields of psychology showed a ratio of 1,675/1,383 for women to men. The percentage increase has been greatest in clinical (46.4), counseling (183.7), and business (253.3).[16]

CLINICAL PSYCHOLOGY IN THE NEAR FUTURE

Because professional interests have come to predominate in the activities and policies of the APA, many of those members dedicated to the values of science and teaching reluctantly concluded that reform within the Association was impossible and that the time had come to form a new organization which would take up the functions that had been so eroded within the APA. Thus, the American Psychological Society (APS) came into being, with a first annual conference in 1988 and an initial membership of several thousand. Some clinical and other psychologists with serious research interests in behavior pathology formed the society for Research in Psychopathology as a forum for the presentation and discussion of scientific contributions to psychopathology. The first annual conference of this organization was held at Harvard in 1986, the fourth (1989) meeting being held at the University of Miami. At the time of this writing a new organization, the American Association of Applied and Preventive Psychology, is in the process of formation intended to support and preserve the original purposes of the field of clinical psychology.

For the time being at least, it seems clear that the scientific interests of the clinical psychologist are likely to be focused in the main in organizations outside the APA. Indeed, the proliferation of specialized societies for various subdisciplines of psychology in general has increased steadily in recent decades and appears to reflect the perception of scientific psychologists that their interests will best be served outside the APA. The interests of the practitioner seem to be solidly established within the APA.

With regard to the practice of psychotherapy, it is possible that economic forces will reduce the number of psychiatrists involved in this, leaving it largely to clinical psychologists and an array of other mental health professionals. Economic forces include the increased competition for what is not a bottomless fund of potential client payments, but more to the point, the fact that psychotherapy is hourly paid. As medical financial policies are such that payment for a procedure is the norm (a simple appendectomy generally costing a certain amount regardless of the time taken to do it), the economics of psychiatric practice provide a disincentive toward hourly payment and a positive incentive to the use of procedures such as the prescription of medication. As psychiatry moves increasingly toward a biological approach to psychopathology, this trend is supported by its effect in bringing psychiatry closer to its medical heritage and further from the psychosocial influences that were predominant in the 1930s to 1960s. Recent advances in biobehavioral measurement, exemplified by the development of brain scan methods, and the rapid expansion of neuropsychology testify to the emphases that we may expect to see in the immediate future.

The events covered in this personal history of clinical psychology have not altered the fact that there are many human beings with serious psychological problems and that their effective treatment still awaits the accumulation of greater knowledge. To gain that knowledge requires a continued and arduous effort, and the task remains as one of the most important to face the clinical psychologist.

REFERENCES

Albee, G. W. (1970). The uncertain future of clinical psychology. *The American Psychologist, 25,* 1071–1080.

[16]Figures from the National Science Foundation, as reported in the *APS Observer,* 1989, *2,* p. 7.

Allyon, T., & Houghton, E. (1962). Control of the behavior of schizophrenic patients by food. *Journal of the Experimental Analysis of Behavior, 5,* 343–352.

Allyon, T., & Michael, J. (1959). The psychiatric nurse as a behavioral engineer. *Journal of the Experimental Analysis of Behavior, 2,* 323–336.

Bennett, C. C., Anderson, L. S., Cooper, S., Hassol, L., Klein, D. L., & Rosenblum, G. (Eds.). (1967). *Community psychology.* Boston: Boston University Press.

Dollard, J., & Miller, N. E. (1950). *Personality and psychotherapy.* New York: McGraw-Hill.

Eysenck, H. J. (1952). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology, 16,* 319–324.

Garfield, S. L., & Kurtz, R. M. (1974). A survey of clinical psychologists: Characteristics, activities and orientations. *The Clinical Psychologist, 28* (1), 7–10.

Goffman, E. (1961). *Asylums.* New York: Anchor-Doubleday.

Hilgard, E. R. (1987). *Psychology in America: A historical survey.* San Diego, CA: Harcourt Brace Jovanovich.

Institute of Medicine. (1983). *Personnel needs and training for biomedical and behavioral research.* Washington, DC: National Academy Press.

Kendall, P. C., & Norton-Ford, J. D. (1982). *Clinical psychology: Scientific and professional dimensions.* New York: John Wiley & Sons.

Kirsch, I., & Winter, C. (1983). A history of clinical psychology. In C. E. Walker (Ed.), *The handbook of clinical psychology: Theory, research and practice* (Vol. 1, pp. 3–30). Homewood, IL. Dow-Jones Irwin.

Korchin, S. J. (1983). The history of clinical psychology: A personal view. In M. Hersen, A. E. Kazdin, & A. S. Bellack (Eds.), *The clinical psychology handbook* (pp. 5–19). Elmsford, NY: Pergamon Press.

Laing, R. D. (1960). *The divided self.* London: Tavistock.

MacCorquodale, K., & Meehl, P. E. (1948). On the distinction between hypothetical constructs and intervening variables. *Psychological Review, 85,* 99–107.

Maher, B. A. (1965). The clinician as a research-psychopathologist. *Preconference materials: Conference on professional preparaton of clinical psychologists.* Washington, DC: American Psychological Association.

Maher, B. A. (1966). *Principles of psychopathology: An experimental approach.* New York: McGraw-Hill.

Maher, B. A. (1970). *Introduction to research in psychopathology.* New York: McGraw-Hill.

Maher, B. A. (1976). *Journal of Clinical and Consulting Psychology,* Editor's Annual Report to the American Psychological Association. Unpublished document.

Maher, B. A. (1982). Mandatory insurance coverage for psychotherapy: A tax on the subscriber and a subsidy to the practitioner. *The Clinical Psychologist, 35* (2), 9–12.

Maher, B. A. (1988). Abnormal and clinical psychology. In E. R. Hilgard (Ed.), *Fifty years of psychology: Essays in honor of Floyd Ruch* (pp. 153–168). Glenview, IL: Scott Foresman.

Maher, B. A., & Maher, W. B. (1979). Psychopathology. In E. Hearst (Ed.), *The first century of experimental psychology* (pp. 561–621). Hillsdale, NJ: Lawrence Erlbaum Associates.

Maher, B. A., & Maher, W. B. (1985). Psychopathology: II. From the eighteenth century to modern times. In G. A. Kimble & K. Schlesinger (Eds.), *Topics in the history of psychology* (pp. 295–329). Hillsdale, NJ: Lawrence Erlbaum Associates.

Maher, W. B., & Maher, B. A. (1985). Psychopathology: I. From ancient times to the eighteenth century. In G. A. Kimble & K. Schlesinger (Eds.), *Topics in the history of psychology* (pp. 251–294). Hillsdale, NJ: Lawrence Erlbaum Associates.

Matarazzo, J. D. (1985). Psychotherapy. In G. A. Kimble & K. Schlesinger (Eds.), *Topics in the history of psychology* (pp. 219–250). Hillsdale, NJ: Lawrence Erlbaum Associates.

McGuire, T. C. (1981). Compulsory insurance for psychotherapy. *The Clinical Psychologist, 34,* 13–14.

McNett, I. (1982). Psy.D. fills the demand for practitioners. *APA Monitor, 13,* 10–11.

Mettler, F. A. (Ed.). (1949). *Selective partial ablation of the frontal cortex.* New York: Hoeber-Harper.

Mettler, F. A. (Ed.). (1952). *Psychosurgical problems.* New York: Blakiston.

Miller, N. E. (1948). Theory and experiment relating psychoanalytic displacement to stimulus response generalization. *Journal of Abnormal and Social Psychology, 43,* 155–178.

Murray, E. J., & Berkun, M. M. (1955). Displacement as a function of conflict. *Journal of Abnormal and Social Psychology, 51,* 47–56.

Norcross, J. C., & Prochaska, J. O. (1982). A national survey of clinical psychologists: Characteristics and activities. *The Clinical Psychologist, 35* (2), 1–8.

Reed, M. R. (1982). Letter to the editor. *The Clinical Psychologist, 36,* 16–17.

Rosenhan, D. L. (1973). On being sane in insane places. *Science, 179,* 250–258.

Szasz, T. (1961). *The myth of mental illness.* New York: Hoeber-Harper.

Wheeler, S. (1968). *Controlling delinquents.* New York: John Wiley & Sons.

Wolpe, J. (1958). *Psychotherapy by reciprocal inhibition.* Palo Alto: Stanford University Press.

CHAPTER 1

HISTORICAL PERSPECTIVES

Gerald Goldstein
Michel Hersen

INTRODUCTION

"A test is a systematic procedure for comparing the behavior of two or more persons." This definition of a test, offered by Lee Cronbach (1949, 1960) many years ago probably still stands as a means of epitomizing the major content of psychological assessment. The invention of psychological tests, then known as mental tests, is generally attributed to Galton (Boring, 1950) and occurred during the middle and late nineteenth century. Galton's work was largely concerned with individual differences, and his approach was essentially the opposite of what was the general case for psychologists of his time. Psychologists then were largely concerned with the exhaustive study of mental phenomena in a few subjects, while Galton was more interested in somewhat less specific analyses of large numbers of people. Perhaps the first psychological test was the "Galton whistle," which evaluated high tone hearing. Galton also appeared to have the statistical concept that errors of measurement in individuals could be canceled out through the mass effect of large samples.

Obviously, psychologists have come a long way from the simple tests of Galton, Binet, and Munsterberg, and the technology of testing has already entered the computer age, with almost science fictionlike extensions, such as testing by satellite. Psychometrics is now an advanced branch of mathematical and statistical science, and the administration, scoring, and even interpretation of tests have become increasingly objectified and automated. While some greet the news with dread and others with enthusiasm, we may be rapidly approaching the day when most or all testing will be administered, scored, and interpreted by computer. Thus, the nineteenth-century image of the school teacher administering paper-and-pencil tests to the students in her classroom and grading them at home has changed to the extensive use of automated procedures administered to huge portions of the population by representatives of giant corporations. Testing appears to have become a part of Western culture, and there are indeed very few people who enter educational, work, or clinical settings who do not take many tests during their lifetimes.

In recent years, there appears to have been a distinction made between testing and assessment, assessment being the broader concept. Psychologists do not just give tests now; they perform assessments. The title of this volume, the *Handbook of Psychological Assessment*, was chosen advisedly and is meant to convey the view that it is not simply a handbook of psychological testing, although testing will be covered in great detail. The term assessment implies that there are many ways of evaluating individual differences. Testing is one way, but there are also interviewing, observations of behavior in natural or structured settings, and recording of various physiological functions. Certain forms of interviewing and systematic observation of behavior are now known as behavioral assessments, as opposed to the psychometric assessment accomplished through the use of formal tests. Historically, interest in these two forms of assessment has waxed and waned, and in what follows we will briefly try to trace these trends in various areas.

INTELLIGENCE TESTING

The testing of intelligence in school children was probably the first major occupation of clinical psychology. The Binet scales and their descendants continue to be used, along with the IQ concept associated with them. Later, primarily through the work of David Wechsler and associates (Wechsler, 1944), intelligence testing was extended to adults and the IQ concept was changed from the mental age system (Mental Age/Chronological Age × 100) to the notion of a deviation IQ based on established norms. While Wechsler was primarily concerned with the individual assessment of intelligence, many group-administered paper-and-pencil tests also emerged during the early years of the twentieth century. Perhaps the old Army Alpha and Beta tests, developed for intellectual screening of inductees into the armed forces during the First World War, were the first examples of these instruments. The use of these tests progressed in parallel with developments in more theoretical research regarding the nature of intelligence. The English investigators, Burt, Pearson, and Spearman, and the Americans, Thurstone and Guilford, are widely known for their work in this area, particularly with factor analysis. The debate over whether intelligence is a general ability (g) or a series of specific abilities represents one of the classic controversies in psychology. A related controversy that is still very much with us (Jensen, 1983) has to do with whether intelligence is primarily inherited or acquired and with the corollary issue having to do with ethnic-racial differences in intellectual ability.

Another highly significant aspect of intelligence testing has to do with its clinical utilization. The IQ now essentially defines the borders of mental retardation, and intelligence tests are widely used to identify retarded children in educational settings (DSM-III [American Psychiatric Association, 1980]). However, intelligence testing has gone far beyond the attempt to identify mentally retarded individuals and has become widely applied in the fields of psychopathology and neuropsychology. With regard to psychopathology, under the original impetus of David Rapaport and collaborators (1945), the Wechsler scales became clinical instruments used in conjunction with other tests to evaluate patients with such conditions as schizophrenia and various stress-related disorders. In the field of neuropsychology, the use of intelligence testing is perhaps best described by McFie's (1975) remark, "It is perhaps a matter of luck that many of the Wechsler subtests are neurologically relevant" (p. 14). In these applications, the intelligence test was basically used as an instrument with which the clinician could examine various cognitive processes, on the basis of which inferences could be made about the patient's clinical status.

In summary, the intelligence test has become a widely used assessment instrument in educational, industrial, military, and clinical settings. While in some applications the emphasis remains on the simple obtaining of a numerical IQ value, it would probably be fair to say that many, if not most, psychologists now use the intelligence test as a means of examining the individual's cognitive processes, of seeing how he or she goes about solving problems, of identifying those factors that may be interfering with adaptive thinking, of looking at various language and nonverbal abilities in brain-damaged patients, and of identifying patterns of abnormal thought processes seen in schizophrenic patients. Performance profiles and qualitative characteristics of individual responses to items appear to have become the major focuses of interest, rather than the single IQ score.

PERSONALITY ASSESSMENT

Personality assessment has come to rival intelligence testing as a task performed by psychologists. However, while most psychologists would agree that an intelligence test is generally the best way to measure intelligence, no such consensus obtains within the realm of personality evaluation. In long-term perspective, it would appear that two major philosophies have emerged and perhaps three assessment methods. The two philosophies can be traced back to Allport's (1937) distinction between nomothetic versus idiographic methodologies and Meehl's (1954) distinction between clinical and statistical or actuarial prediction. In essence, some psychologists feel that personality assessments are best accomplished when they are highly individualized, while others have a preference for quantitative procedures based on group norms. The phrase "seer versus sign" has been used to epitomize this dispute. The three methods referred to are the interview, the projective, and the objective tests. Obviously, the first way psychologists and their predecessors found out about people was to talk to them, giving the interview historical precedence. But following a period when the use of the interview was eschewed by many psychologists, it has made a return. It would appear that the field is in a historical spiral, with various methods leaving and returning at different levels.

The interview began as a relatively unstructured conversation with the patient and perhaps an informant, with varying goals, including obtaining a history, assessing personality structure and dynamics, establishing a diagnosis, and many other issues. Numerous publications have been written about interviewing (e.g., Menninger, 1952), but in general they provided outlines and general guidelines as to what should be accomplished by the interview. However, model interviews were not provided. With or without this guidance, the interview was viewed by many as a subjective, unreliable procedure that could not be sufficiently validated. For example, the unreliability of psychiatric diagnosis based on studies of multiple interviewers had been well established (Zubin, 1967). More recently, however, a number of structured psychiatric interviews have appeared in which the specific content, if not specific items, has been presented, and for which very adequate reliability has been established. There are by now several such interviews available, including the Schedule for Affective Disorders and Schizophrenia (SADS) (Spitzer & Endicott, 1977), the Renard Diagnostic Interview (Helzer, Robins, Croughan, & Welner, 1981) and the Structured Clinical Interview for DSM-III or DSM-III-R (SCID or SCID-R) (Spitzer & Williams, 1983). These interviews have been established in conjunction with objective diagnostic criteria including DSM-III itself, the Research Diagnostic Criteria (Spitzer, Endicott, & Robins, 1977), and the Feighner Criteria (Feighner, Robins, Guze, Woodruff, Winokur, & Munoz, 1972). These new procedures have apparently ushered in a "comeback" of the interview, and many psychiatrists and psychologists now prefer to use these procedures rather than either the objective- or projective-type psychological test.

Those advocating use of structured interviews note that in psychiatry, at least, tests ultimately must be validated against judgments made by psychiatrists. These judgments are generally based on interviews and observation, because there really are no biological or other objective markers of most forms of psychopathology. If that is indeed the case, there seems little point in administering elaborate and often lengthy tests when one can just as well use the criterion measure itself, the interview, rather than the test. There is no way that a test can be more valid than an interview if an interview is the validating criterion. The structured interviews have made a major impact on the scientific literature in psychopathology, and it is rare to find a recently written research report in which diagnoses were not established by one of them. It would appear that we have come full cycle regarding this matter, and until objective markers of various forms of psychopathology are discovered, we will be relying primarily on the structured interviews for our diagnostic assessments.

Interviews of the SADS or Diagnostic Interview Schedule type are relatively lengthy and comprehensive, but there are now several briefer and more specific interview or interview-like procedures. Within psychiatry, perhaps the most well-known procedure is the Brief Psychiatric Rating Scale (BPRS) (Overall & Gorham, 1962). The BPRS is a brief, structured, repeatable interview that has essentially become the standard instrument for assessment of change in patients, usually as a function of taking some form of psychotropic medication. In the specific area of depression, the Hamilton Depression Scale (Hamilton, 1960) plays a similar role. There are also several widely used interviews for patients with dementia, which generally combine a brief mental status examination with some form of functional assessment, particularly with reference to Activities of Daily Living. The most popular of these scales are the Mini-Mental Status Examination of Folstein, Folstein, and McHugh (1975) and the Dementia Scale of Blessed, Tomlinson, and Roth (1968). Extensive validation studies have been conducted with these instruments, perhaps the most well-known one having to do with the correlation between scores on the Blessed, Tomlinson, and Roth scale during life and the senile plaque count determined on autopsy in patients with dementia. The obtained correlation of .7 quite impressively suggested that the scale was a valid one for detection of dementia. In addition to these interviews and rating scales, numerous methods have been developed for assessment of psychopathology by nurses and psychiatric aides based on direct observation of ward behavior (Raskin, 1982). The most widely used of these rating scales are the Nurses' Observation Scale for Inpatient Evaluation (NOSIE-30) (Honigfeld & Klett, 1965) and the Ward Behavior Inventory (Burdock, Hardesty, Hakerem, Zubin, & Beck, 1968). These scales assess such behaviors as cooperativeness, appearance, communication, aggressive episodes, and related behaviors, and are to be based on direct observation rather than reference to medical records or the report of others. Scales of this type supplement the interview with information concerning social competence and capacity to carry out functional activities of daily living.

Again taking a long-term historical view, it is our impression that after many years of neglect by the field, the interview has made a successful return to the arena of psychological assessment, but the interviews

now used are quite different from the loosely organized, "free-wheeling," conversation-like procedures of the past. First of all, their organization tends to be structured, and the interviewer is required to obtain certain items of information. It is generally felt that formulation of specifically worded questions is counterproductive; rather, the interviewer, who should be an experienced clinician trained in the use of the procedure, should be able to formulate questions that will elicit the required information. Second, the interview procedure must meet psychometric standards of reliability and validity. Finally, while the structured interviews tend to be atheoretical in orientation, they are based on contemporary scientific knowledge of psychopathology. Thus, for example, the information needed to establish a differential diagnosis within the general classification of the affective disorders is derived from the scientific literature on depression and the related affective disorders.

The rise of the interview appears to have occurred in parallel with the decline of projective techniques. Those of us now in a chronological category that may be roughly described as middle-aged may recall that their graduate training in clinical psychology probably included extensive course work and practicum experience involving the various projective techniques. Most clinical psychologists would probably agree that even though projective techniques are still used to some extent, the atmosphere of ferment and excitement concerning these procedures that existed during the 1940s and 1950s no longer seems to exist. Even though the Rorschach technique and Thematic Apperception Test (TAT) were the major procedures used during that era, a variety of other tests emerged quite rapidly: the projective use of human figure drawings (Machover, 1949), the Szondi Test (Szondi, 1952), the Make-A-Picture-Story (MAPS) Test (Shneidman, 1952), the Four Picture Test (VanLennep, 1951), the Sentence Completion Tests (e.g., Rohde, 1957), and the Holtzman Inkblot Test (Holtzman, 1958). The exciting work of Murray and his collaborators reported on in *Explorations in Personality* (Murray, 1938) had a major impact on the field and stimulated extensive utilization of the TAT. It would probably be fair to say that the sole survivor of this active movement is the Rorschach. Many clinicians continue to use the Rorschach, and the work of Exner and his collaborators has lent it increasing scientific respectability (see Erdberg, chapter 17).

There are undoubtedly many reasons for the decline in utilization of the projective techniques, but in our view they can be summarized by the following points.

1. Increasing scientific sophistication created an atmosphere of skepticism concerning these instruments. Their reliability and validity were called into question by numerous studies (e.g., Swensen, 1957, 1968; Zubin, 1967), and it was felt by a substantial segment of the professional community that the claims made for these procedures could not be substantiated.
2. Developments in alternative procedures, notably the MMPI and other objective tests, convinced many clinicians that the information previously adduced from projective tests could be gained more efficiently and less expensively with the objective methods. In particular, the voluminous MMPI research literature has demonstrated its usefulness in an extremely wide variety of clinical and research settings. When the MMPI and related objective techniques were pitted against the projectives during the days of the "seer versus sign" controversy, it was generally demonstrated that the sign was as good as or better than the seer in most of the studies accomplished (Meehl, 1954).
3. In general, the projective techniques are not atheoretical and, in fact, are generally viewed as being associated with one or another branch of psychoanalytic theory.

While psychoanalysis remains a strong and vigorous movement within psychology, there are numerous alternative theoretical systems at large, notably behaviorally and biologically oriented systems. As is implied in the section of this chapter covering behavioral assessment, behaviorally oriented psychologists pose theoretical objections to the projective techniques and make little use of them in their practices. Similarly, the projective techniques tend not to receive high levels of acceptance in biologically oriented psychiatry departments. In effect, then, utilization of the projective techniques declined for scientific, practical, and philosophical reasons. However, the Rorschach, in particular, continues to be productively used, primarily by psychodynamically oriented clinicians.

The early history of the objective personality tests has been traced by Cronbach (1949, 1960). The beginnings apparently go back to Sir Francis Galton, who devised personality questionnaires during the latter part of the nineteenth century. We will not repeat that history here, but rather will focus on those procedures that survived into the contemporary era. In our view, there have been three such major survivors: a series of tests developed by Guilford and collaborators (Guilford & Zimmerman, 1949), a similar series developed by Cattell and collaborators (Cattell, Eber, & Tatsuoka, 1970), and the MMPI. In general, but certainly not in all cases, the Guilford and Cattell procedures are used for individuals functioning within the normal range, while the MMPI is more widely used in clinical populations. Thus, for example, Cattell's 16PF test may be used to screen job applicants while the MMPI may be more typically used in psychiatric health care facilities. Furthermore, the Guilford and Cattell tests are based on factor analysis and are trait oriented, while the MMPI in its standard form does not make use of factor analytically derived scales and is more oriented toward psychiatric classification. Thus, the Guilford and Cattell scales contain measures of such traits as dominance or sociability, while most of the MMPI scales are named after psychiatric classifications such as paranoia or hypochondriasis.

Currently, most psychologists use one or more of these objective tests rather than interviews or projective tests in screening situations. For example, many thousands of patients admitted to Veterans Administration psychiatric facilities take the MMPI shortly after admission, while applicants for prison guard jobs in the state of Pennsylvania take the Cattell 16PF. However, the MMPI in particular is commonly used as more than a screening instrument. It is frequently used as a part of an extensive diagnostic appraisal, as a method of evaluating treatment, and in numerous research applications. There is little question that it is the most widely used and extensively studied procedure in the objective personality test area. Even though the 566 true or false items in the first edition have remained the same since the initial development of the instrument, the test's applications and clinical interpretation have evolved dramatically over the years. We have gone from a perhaps overly naive dependence on single-scale evaluations and overly literal interpretation of the names of the scales (many of which are archaic psychiatric terms) to sophisticated configural interpretation of profiles, much of which is based on empirical research (Gilberstadt & Duker, 1965; Marks, Seeman, & Haller, 1974). Correspondingly, the methods of administering, scoring, and interpreting the MMPI have kept pace with technological and scientific advances in the behavioral sciences. Beginning with sorting cards into piles, hand scoring, and subjective interpretation, the MMPI has gone to computerized administration and scoring, interpretation based at least to some extent on empirical findings, and even computerized interpretation (Fowler, 1979, and chapter 23, this volume). As is well known, there are several companies that will provide computerized scoring and interpretations of the MMPI.

Since the appearance of the first edition of this handbook, there have been two major developments in the field of objective personality assessment. First, Millon has produced a series of tests referred to as The Millon Clinical Multiaxial Inventory (Versions I and II), The Millon Adolescent Personality Inventory and the Millon Behavioral Health Inventory (Millon, 1982, 1985). Second, the MMPI has been completely revised and restandardized, and now is available as the MMPI-2. Chapter 16 in this volume by Keller, Butcher, and Slutske describes these new developments in detail.

Even though we should anticipate continued spiraling of trends in personality assessment, it would appear that we have passed an era of projective techniques and are now living in a time of objective assessment with an increasing interest in the structured interview. There also appears to be increased concern with the scientific status of our assessment procedures. In recent years there has been particular concern with reliability of diagnosis, especially since distressing findings appeared in the literature suggesting that psychiatric diagnoses were being made quite unreliably (Zubin, 1967). The issue of validity in personality assessment, however, remains a difficult one for a number of reasons. First, if by personality assessment we mean prediction or classification of some psychiatric diagnostic category, we have the problem of there being essentially no known objective markers for the major forms of psychopathology. Therefore, we are left essentially with psychiatrists' judgments. DSM-III and its revision have greatly improved this situation by providing objective criteria for the various mental disorders, but the capacity of such instruments as the MMPI or Rorschach to predict DSM-III and DSM-III-R diagnoses has not yet been evaluated and remains a research question for the future. Some scholars, however, even question the usefulness of taking that research course rather than developing increasingly reliable and valid structured interviews (Zubin, 1984). Similarly, there have been many reports of the failure of objective tests to predict such matters as success in an occupation or trustworthiness with regard to handling a weapon. For example, objective tests are no longer used to screen astronauts, since they were not successful in predicting who would be successful or unsuccessful (Cordes, 1983). There does, in fact, appear to be a movement

within the general public and the profession toward discontinuation of use of personality assessment procedures for decision making in employment situations. We would note as another possibly significant trend a movement toward direct observation of behavior in the form of behavioral assessment. The zeitgeist definitely is in opposition to procedures in which the intent is disguised. Burdock and Zubin (1985), for example, argue that "nothing has as yet replaced behavior for evaluation of mental patients."

NEUROPSYCHOLOGICAL ASSESSMENT

Another area that has an interesting historical development is neuropsychological assessment. The term itself is a relatively new one and probably was made popular through the first edition of Lezak's (1976) book with that title. Neuropsychological assessment is of particular historical interest because it represents a confluence of two quite separate antecedents: central and eastern European behavioral neurology, and American and English psychometrics. Neurologists, of course, have always been concerned with the behavioral manifestations of structural brain damage and the relationship between brain function and behavior. Broca's discovery of a speech center in the left frontal zone of the brain is often cited as the first scientific neuropsychological discovery, because it delineated a relatively specific relationship between a behavioral function, speech, and a correspondingly specific region of the brain (the third frontal convolution of the left hemisphere). Clinical psychologists developed an interest in this area when they were called upon to assess patients with known or suspected brain damage. The first approach to this diagnostic area involved utilization of the already existing psychological tests, and the old literature in the area deals primarily with how such tests as the Wechsler scales, the Rorschach, or the Bender-Gestalt test could be used to diagnose brain damage. More recently, special tests were devised specifically for assessment work with patients having known or suspected brain damage.

The merger between clinical psychology and behavioral neurology can be said to have occurred when the sophistication of neurologists relative to the areas of brain function and brain disease was combined with the psychometric sophistication of clinical psychology. The wedding occurred when reliable, valid, and well-standardized measurement instruments began to be used to answer complex questions in neurological diagnosis and differential diagnosis. Thus, clinicians who ultimately identified themselves as clinical neuropsychologists tended to be individuals who knew their psychometrics, but who also had extensive training and experience in neurological settings. Just as many clinical psychologists work with psychiatrists, many clinical neuropsychologists work with neurologists and neurosurgeons. This relationship culminated in the development of standard neuropsychological test batteries, notably the Halstead-Reitan (Reitan & Wolfson, 1985) and Luria-Nebraska batteries (Golden, Hammeke, & Purisch, 1980), as well as in the capacity of many trained psychologists to perform individualized neuropsychological assessments of adults and children. Thus, within the history of psychological assessment, clinical neuropsychological evaluation has recently emerged as an independent discipline to be distinguished from general clinical psychology on the basis of the specific expertise members of that discipline have in the areas of brain-behavior relationships and diseases of the nervous system. In recent years, there have been expansions of both the standard batteries and the individual neuropsychological tests. An alternate form (Golden, Purisch, & Hammeke, 1985) as well as a children's version (Golden, 1981) of the Luria–Nebraska Neuropsychological Battery are now available. Prominent among the newly published or revised individual tests are the series of tests described in detail by Arthur Benton and collaborators in *Contributions to Neuropsychological Assessment* (Benton, Hamsher, Varney, & Spreen, 1983), the California Verbal Learning Test (Delis, Kramer, Kaplan, & Ober, 1987), and the newly revised and thoroughly reworked Wechsler Memory Scale (WMS-R) (Wechsler, 1987).

BEHAVIORAL ASSESSMENT

Over the last three decades behavioral assessment has been one of the most exciting developments to emerge in the field of psychological evaluation (Bellack & Hersen, 1988). Although its seeds were planted long before behavior therapy became a popular therapeutic movement, it is with the advent of behavior therapy that the strategies of behavioral assessment began to flourish (cf. Hersen & Bellack, 1976, 1981). As has been noted elsewhere (Barlow & Hersen, 1984; Herson & Barlow, 1976), behavioral assessment can be conceptualized as a reaction to a number of factors. Among these were the (a) problems with unreliability and invalidity of aspects of the DSM-I and DSM-II diagnostic schemes, (b) concerns over the indirect relationship between what was evaluated in traditional testing (e.g., the projectives) and how it subsequently was used in treatment planning and application, (c) increasing acceptance of behavior therapy by the professional community as a viable series of therapeutic modalities, and (d) parallel developments in the field of diagnosis in general involving greater precision and accountability (e.g., the problem-oriented record).

We will briefly consider each of the four factors in turn and see how they contributed historically to the development of behavioral assessment. To begin with, DSM-I and DSM-II were the targets of considerable criticism from psychiatrists (Hines & Williams, 1975) and psychologists alike (Begelman, 1975). Indeed, Begelman (1975), in a humorous vein, referred to the two systems as "twice-told tales." They were "twice told" in the sense that neither resulted in highly reliable classification schemes when patients were independently evaluated by separate psychiatric interviewers (cf. Ash, 1949; Sandifer, Pettus, & Quade, 1964). Problems were especially evident when attempts to obtain interrater reliability were made for the more minor diagnostic groupings of the DSM schemes. Frequently, clinical psychologists would be consulted to carry out their testing procedures to confirm or disconfirm psychiatrists' diagnostic impressions based on DSM-I and DSM-II. But in so doing, such psychologists, operating very much as x-ray technicians, were using procedures (objective and projective tests) that had only a tangential relationship to the psychiatric descriptors for each of the nosological groups of interest. Thus, the futility of this kind of assessment strategy over time became increasingly apparent. Moreover, not only were there problems with the reliability for DSM-I and DSM-II, but empirical studies documented considerable problems as well with regard to external validity of the systems (Eisler & Polak, 1971: Nathan, Zare, Simpson, & Ardberg, 1969).

Probably more important than any of the above was the fact that the complicated psychological evaluation had a limited relationship to eventual treatment. At least in the psychiatric arena, the usual isomorphic relationship between assessment and treatment found in other branches of therapeutics did not seem to hold. The isolated and extended psychological examination frequently proved to be an empty academic exercise resulting in poetic jargon in the report that eventuated. But its practical utility was woefully limited. Treatment seemed to be unrelated to the findings in the reports.

All of the aforementioned resulted in attempts by clinical psychologists to measure the behaviors of interest in direct fashion. For example, if a patient presented with a particular phobia, the objective of evaluation was not to assess the underlying "neurotic complex" or "alleged psychodynamics." Quite the contrary, the primary objective was to quantify in distance how close our patient could approach the phobic object (i.e., the behavioral approach task) and how his heart rate (physiological assessment) increased as he got closer. In addition, the patient's cognitions (self-report) were quantified by having him assess his level of fear (e.g., on a 1–10 point scale). Thus, the behavioral assessment triad, consisting of motoric, physiological, and self-report systems (Hersen, 1973), was established as the alternative to indirect measurement.

Commenting on the use of direct measurement, Hersen and Barlow (1976) argue that

> whereas in indirect measurement a particular response is interpreted in terms of a presumed underlying disposition, a response obtained through direct measurement is simply viewed as a sample of a large population of similar responses elicited under those particular stimulus conditions. . . . Thus, it is hardly surprising that proponents of direct measurement favor the observation of individuals in their natural surroundigs whenever possible. When such naturalistic observations are not feasible, analogue situations approximating naturalistic conditions may be developed to study the behavior in question (e.g., the use of a behavioral avoidance test to study the degree of fear of snakes). When neither of these two methods is available or possible, subjects' *self-reports are also used as independent criteria,* and, at times, may be operating under the control of totally different sets of contingencies than those governing motoric responses. (p. 116)

We already have referred to the tripartite system of direct measurement favored by the behaviorists. But it is in the realm of motoric behavior that behavior therapists have made the greatest contributions as well as being most innovative (see Foster, Bell-Dolan, & Burge, 1988; Hersen, 1988; Tryon, 1986). With increased acceptance of behavior therapy, practitioners of the strategies found their services required in a large variety of educational, rehabilitation, community medical, and psychiatric settings. Very often they were presented with extremely difficult educational, rehabilitation, and treatment cases, both from assessment and therapeutic perspectives. Many of the clients and patients requiring remediation exhibited behaviors that previously had not been measured in any direct fashion. Thus, there were few guidelines with regard to how the behavior might be observed, quantified, and coded. In many instances, "seat-of-the-pants" measurement systems were devised on the spot

but with little regard for psychometric qualities cherished by traditional testers.

Consider the following example of a measurement strategy to quantify "spasmodic torticollis," a tic-like disorder (Bernhardt, Hersen, & Barlow, 1972):

> A Sony video recorder model AV-5000A, an MRI Keleket model VC-1 television camera, and a Conrac 14-inch television monitor were employed in recording torticollis. A Gra Lab sixty-minute Universal Timer was used to obtain percentage of torticollis. . . . A lightolier lamp served as the source of negative feedback. Two to three daily ten-minute sessions were scheduled during the experiment in which the subject was videotaped while seated in a profile arrangement. A piece of clear plastic containing superimposed Chart-Pac taped horizontal lines (spaced one-quarter to one-half inch apart) was placed over the monitor. A shielded observer depressed a switch activating the timer whenever the subject's head was positioned at an angle where the nostril was above a horizontal line intersecting the external auditory meatus. This position was operationally defined as an example of torticollis, with percentage of torticollis per session serving as the experimental measure. Conversely, when the horizontal line intersected both the nostril and auditory meatus or when the subject's nostril was below the horizontal line he was considered to be holding his head in a normal position. (p. 295)

If one peruses the pages of the *Journal of Applied Behavior Analysis, Behaviour Research and Therapy, Journal of Behavior Therapy and Experimental Psychiatry,* and *Behavior Modification,* particularly in the earlier issues, numerous examples of innovative behavioral measures and more comprehensive systems are to be found. Consistent with the idiographic approach, many of these apply only to the case in question, have some internal or face validity, but, of course, have little generality or external validity. (We will further comment on this aspect of behavioral assessment in a subsequent section of this chapter.)

A final development that contributed to and coincided with the emergence of behavioral assessment was the problem-oriented record (POR). This was a system of recordkeeping first instituted on medical wards in general hospitals to sharpen and pinpoint diagnostic practices (cf. Weed, 1964, 1968, 1969). Later this system was transferred to psychiatric units (cf. Hayes-Roth, Longabaugh, & Ryback, 1972; Katz & Woolley, 1975; Klonoff & Cox, 1975; McLean & Miles, 1974; Scales & Johnson, 1975), with its relevance to behavioral assessment increasingly evident (Atkinson, 1973; Katz & Woolley, 1975). When applied to psychiatry, the POR can be divided into four sections: (a) data base, (b) problem list, (c) treatment plan, and (d) follow-up data. There can be no doubt that this kind of recordkeeping promotes and enhances the relationship of assessment and treatment, essentially forcing the evaluator to crystallize his or her thinking about the diagnostic issues. In this regard, we previously have pointed out that

> despite the fact that POR represents, for psychiatry, a vast improvement over the type of record-keeping and diagnostic practice previously followed, the level of precision in describing problem behaviors and treatments to be used remedially *does not* yet approach the kind of precision reached in the carefully conducted behavioral analysis. (Hersen, 1976, p. 15)

However, the POR certainly can be conceptualized as a major step in the right direction. In most psychiatric settings some type of POR is currently being used and, to a large extent, has further legitimized the tenets of behavioral assessment by clearly linking the problem list with specific treatment (cf. Longabaugh, Fowler, Stout, & Kriebel, 1983; Longabaugh, Stout, Kriebel, McCullough, & Bishop, 1986).

ASSESSMENT SCHEMES

Since 1968 a number of comprehensive assessment schemes have been developed to facilitate the process of behavioral assessment (Cautela, 1968; Kanfer & Saslow, 1969; Lazarus, 1973). Because a very detailed analysis of these schemes is much beyond the scope of this brief historical overview, we will only describe the outlines of each in order to illustrate how the behavioral assessor conceptualizes his or her cases. For example, Cautela (1968) depicted in his scheme the role of behavioral assessment during the various stages of treatment. Specifically, he delineated three stages.

In the *first stage* the clinician identifies maladaptive behaviors and those antecedent conditions maintaining them. This step is accomplished through interviews, observation, and self-report questionnaires. The *second stage* involves the selection of the appropriate treatment strategies, evaluation of their efficacy, and the decision when to terminate their application. In the *third stage* a meticulous follow-up of treatment outcome is recommended. This is done by examining motoric, physiologic, and cognitive functioning of the client, in addition to independent confirmation of the client's progress by friends, relatives, and employers.

A somewhat more complicated approach to initial evaluation was proposed by Kanfer and Saslow (1969), which involves some seven steps. The *first* involves a determination as to whether a given behavior represents an excess, deficit, or an asset. The *second* is a clarification of the problem and is based on the notion that in order to be maintained, maladjusted behavior requires continued support. *Third* is the motivational analysis in which reinforcing and aversive stimuli are identified. *Fourth* is the developmental analysis, focusing on biological, sociological, and behavioral changes. *Fifth* involves assessment of self-control and whether it can be used as a strategy during treatment. *Sixth* is the analysis of the client's interpersonal life, and *seventh* is the evaluation of the patient's socio-cultural-physical environment.

In their initial scheme, Kanfer and Saslow (1969) viewed the system in complementary fashion to the existing diagnostic approach (i.e., DSM-II). They did not construe it as supplanting DSM-II, but they did see their seven-part analysis as serving as a basis for arriving at decisions for precise behavioral interventions, thus yielding a more isomorphic relationship between assessment and treatment. Subsequently, Kanfer and Grimm (1977) have turned their attention to how the interview contributes to the overall behavioral assessment. In so doing, suggestions are made for organizing client complaints under five categories:

> (1) behavioral deficiencies, (2) behavioral excesses, (3) inappropriate environmental stimulus control, (4) inappropriate self-generated stimulus control, and (5) problematic reinforcement contingencies. (p. 7)

Yet another behavioral assessment scheme had been proposed by Lazarus (1973), with the somewhat humorous acronym of BASIC ID: B = behavior A = affect, S = sensation, I = imagery, C = cognition, I = interpersonal relationship, and D = the need for pharmacological intervention (i.e., drugs) for some psychiatric patients. The major issue underscored by this diagnostic scheme is that if any of the elements is overlooked, assessment will be incomplete, thus resulting in only a partially effective treatment. To be fully comprehensive, deficits or surpluses for each of the categories need to be identified so that specific treatments can be targeted for each. This, then, should ensure the linear relationship between assessment and treatment, ostensibly absent in the nonbehavioral assessment schemes.

Despite development of the aforementioned schemes and others not outlined here (e.g., Bornstein, Bornstein, & Dawson, 1984), there is little in the way of their formal evaluation in empirical fashion. Although these schemes certainly appear to have a good bit of face validity, few studies, if any, have been devoted to evaluating concurrent and predictive validity. This, of course, is in contrast to the considerable effort to validate the third edition of DSM (i.e., DSM-III, APA, 1980; Hersen & Turner, 1984) and its revision (i.e., DSM-III-R; APA, 1987).

In a somewhat different vein, Wolpe (1977) has expressed his concern about the manner in which behavioral assessment typically is being conducted. Indeed, he has referred to it as the "Achilles' Heel of Outcome Research in Behavior Therapy." He is especially concerned that too little attention has been devoted to evaluation of the antecedents of behaviors targeted for treatment, thus leading to a therapeutic approach that may be inappropriate. For example, in treating anxiety, Wolpe (1977) rightly argues that

> A treatment like systematic desensitization, designed for the deconditioning of anxiety, has very little informational content. It can therefore be expected to have little or no effect on phobic cases that call for cognitive solutions. Such cases should properly be weeded out when the efficacy of desensitization is the topic of research. Until now, as far as I know, such weeding out has not been undertaking in any outcome study involving desensitization. (p. 1)

The same analysis, of course, holds true for disorders, such as depression (Wolpe, 1986) and agoraphobia (Michelson, 1984, 1986). Blanket treatment that does not take into account antecedents undoubtedly should fail (Wolpe & Wright, 1988). But here too, the necessary research findings to document this are as yet forthcoming (see White, Turner, & Turkat, 1983).

CHANGES IN BEHAVIORAL ASSESSMENT

Contrasted to the field of psychological assessment in general, behavioral assessment as a specialty has had a history of about three decades. However, in these years we have witnessed some remarkable changes in the thinking of behavioral assessors. Probably as a strong overt reaction to the problems perceived by behavioral assessors in traditional psychological evaluation, many of the sound psychometric features of that tradition were initially abandoned. Indeed, in some instances it appears that "the baby was thrown out with the bath water." As we already have noted, consistent with the idiographic approach to evaluation and treatment, little concern was accorded to traditional issues of reliability and validity. (The exception, of course, was the obsessive concern with high interrater reliability of observations of motoric behavior.) This was particularly the case for the

numerous self-report inventories developed early on to be consistent with the motoric targets of treatment (e.g., some of the fear survey schedules).

There were many other aspects of traditional evaluation that also were given short shrift. Intelligence testing was eschewed, norms and developmental considerations were virtually ignored, and traditional psychiatric diagnosis was viewed as anathema to behavior therapy. However, since the late 1970s this "hard line" has been mollified. With publication of the second and third editions of *Behavioral Assessment: A Practical Handbook* and emergence of two assessment journals (*Behavioral Assessment* and *Journal of Psychopathology and Behavioral Assessment*), greater attention to cherished psychometric principles has returned. For example, the external validity of role playing as an assessment strategy in the social skill areas has been evaluated by Bellack and his colleagues (cf. Bellack, Hersen, & Lamparski, 1979; Bellack, Hersen, & Turner, 1979; Bellack, Turner, Hersen, & Luber, 1980) instead of being taken on faith. Also in numerous overviews the relevance of the psychometric tradition to behavioral assessment has been articulated with considerable vigor (e.g., Adams & Turner, 1979; Cone, 1977, 1988; Haynes, 1978; Nelson & Hayes, 1979; Rosen, Sussman, Mueser, Lyons, & Davis, 1981). Looking at behavioral assessment today from a historical perspective, it certainly appears as though the "baby" is being returned from the discarded bath water.

Also, in recent years there have been several calls for a broadened conceptualization of behavioral assessment (e.g., Hersen, 1988; Hersen & Bellack, 1988; Hersen & Last, 1989). Such broadening has been most noticeable with respect to the use of intelligence tests in behavioral assessment (Nelson, 1980), the relevance of neuropsychological evaluation for behavioral assessment (Goldstein, 1979; Horton, 1988), the importance of developmental factors especially in child and adolescent behavioral assessment (Edelbrock, 1984; Harris & Ferrari, 1983; Hersen & Last, 1989), and the contribution that behavioral assessment can make to pinpointing of psychiatric diagnosis (Hersen, in press; Tryon, 1986).

DSM-III, DSM-III-R, AND BEHAVIORAL ASSESSMENT

In the earlier days of behavioral assessment, traditional psychiatric diagnosis was for the most part eschewed. Behavioral assessors saw little relationship between what they were doing and the overall implicit goals of DSM-II. Moreover, as we have noted, categories subsumed under DSM-II had major problems with reliability and validity. So, consistent with cogent criticisms about the official diagnostic system, behavioral assessors tended to ignore it when possible. They continued to develop their strategies independently of DSM-II and the then-emerging DSM-III. In fact, some (e.g., Adams, Doster, & Calhoun, 1977; Cautela, 1973) advocated totally new diagnostic formats altogether, but these never had a chance of being accepted by the general diagnostic community, given the political realities.

In spite of its problems and limitations, with the emergence of DSM-III (APA, 1980), behavioral therapists and researchers appear to have retrenched and assumed a somewhat different posture (cf. Hersen, 1988; Hersen & Bellack; 1988; Hersen & Turner, 1984; Nelson, 1987). Such positions have been articulated by a number of prominent behavior therapists, such as Nathan (1981) and Kazdin (1983). But the issues concerning DSM-III and behavioral assessment are most clearly summarized by Taylor (1983), a behavioral psychiatrist:

> The new Diagnostic and Statistical Manual of the American Psychiatric Association is a major improvement in psychiatric diagnosis over previous classification systems. Where symptomatic diagnoses are useful, as in relating an individual's problem to the wealth of clinical and research data in abnormal psychology or in identifying conditions which require specific treatments, DSM-III represents the best available system. Many conceptual and practical problems remain with DSM-III; for instance, it retains a bias toward the medical model, it includes many conditions which should not fall into a psychiatric diagnostic system, and it includes descriptive axes which have not been adequately validated. Nevertheless, behavior therapists are well advised to become familiar with and use DSM-III as part of behavioral assessment. (p. 13)

We, of course, would argue that the same holds true for DSM-III-R and its eventual successor, DSM-IV. We are fully in accord with Taylor's comments and believe that if behavior therapists wish to have an impact on the accepted nosological system, they are urged to work from within rather than from without. In this connection, Tryon (1986) has presented the field with a marvelous outline for how motoric measurements in both children and adults will enable the DSM categories to gain greater precision. He clearly shows how many of the diagnostic categories (e.g., depression; attention deficit-hyperactivity disorders) have motoric referents that could be evaluated by behavioral assessors. However, much work of a normative nature (to determine lower and upper limits of normality) will be required before any impact on the DSM system will be felt (Tryon, 1989). We believe that such evaluation represents an enormous challenge to behavioral assessors that could result in a lasting contribution to the diagostic arena.

SUMMARY

We have attempted to provide a brief historical overview of several major areas in psychological evaluation: intellectual, personality, neuropsychological, and behavioral assessment. Some of these areas have lengthy histories, and others are relatively young. However, it seems clear that the tools used by psychologists as recently as 25 years ago are generally different from those used now. Behavioral assessment techniques, structured psychiatric interviews, and standard, comprehensive neuropsychological test batteries are all relatively new. Furthermore, the computer is at least beginning to make significant inroads into the assessment field, with online testing, scoring, and interpretation a reality in some cases. Serious efforts have been made in recent years to link assessment more closely to treatment and other practical concerns. We may also note a trend away from indirect methods to direct acquisition of information and observation. The structured interview is an example of the former approach, and many behavioral assessment techniques would exemplify the latter one. Similarly, while neuropsychological assessment is still heavily dependent on the use of formal tests, there is increasing interest in the use of those tests in rehabilitation planning and in the association between neuropsychological test results and functional activities of daily living. We also note a corresponding decrease in interest in such matters as brain localization, particularly since the CT scan and related brain imaging procedures have solved much of that problem. We would prognosticate that psychological assessment will be increasingly concerned with automation, the direct observation of behavior, and the practical application of assessment results.

REFERENCES

Adams, H. E., Doster, J. A., & Calhoun, K. S. (1977). A psychologically based system of response classification. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), *Handbook of behavioral assessment*. New York: Wiley.

Adams, H. E., & Turner, S. M. (1979). Editorial. *Journal of Behavioral Assessment, 1,* 1–2.

Allport, G. W. (1937). Personality: *A psychological interpretation*. New York: Holt.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed. rev.). Washington, DC: Author.

Ash, P. (1949). The reliability of psychiatric diagnosis. *Journal of Abnormal and Social Psychology, 44,* 272–276.

Atkinson, C. (1973). Data collection and program evaluation using the problem-oriented medical record. Miami: Association for Advancement of Behavior Therapy.

Barlow, D. H., & Hersen, M. (1984). *Single-case experimental designs: Strategies for studying behavior change* (2nd ed.). Elmsford, NY: Pergamon Press.

Begelman, D. A. (1975). Ethical and legal issues in behavior modification. In M. Hersen, R. M. Eisler, & P. M. Miller (Eds.), *Progress in behavior modification* (Vol. 1). New York: Academic Press.

Benton, A. L., Hamsher, K. deS., Varney, N. R., & Spreen, O. (1983). *Contributions to neuropsychological assessment: A clinical manual*. New York: Oxford University Press.

Bellack, A. S., & Hersen, M. (1988). Future directions. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (3rd ed.). Elmsford, NY: Pergamon Press.

Bellack, A. S., Hersen, M., & Lamparski, D. (1979). Role-playing tests for assessing social skills: Are they valid? Are they useful? *Journal of Consulting and Clinical Psychology, 47,* 335–342.

Bellack, A. S., Hersen, M., & Turner, S. M. (1979). Relationship of role playing and knowledge of appropriate behavior to assertion in the natural environment. *Journal of Consulting and Clinical Psychology, 47,* 679–685.

Bellack, A. S., Turner, S. M., Hersen, M., & Luber, R. (1980). Effects of stress and retesting on role playing tests of social skill. *Journal of Behavioral Assessment, 2,* 99–104.

Bernhardt, A. J., Hersen, M., & Barlow, D. H. (1972). Measurement and modification of spasmodic torticollis: An experiment analysis. *Behavior Therapy, 3,* 294–297.

Blessed, G., Tomlinson, B. E., & Roth, M. (1968). The association between quantitative measures of dementia and of senile change in the cerebral grey

matter of elderly subjects. *British Journal of Psychiatry, 114,* 797–811.
Boring, E. G. (1950). *A history of experimental psychology*. New York: Appleton-Century-Crofts.
Bornstein, P. H., Bornstein, M. T., & Dawson, B. (1984). Integrated assessment and treatment. In T. H. Ollendick & M. Hersen (Eds.), *Child behavioral assessment: Principles and procedures*. Elmsford, NY: Pergamon Press.
Burdock, E. I., Hardesty, A. S., Hakerem, G., Zubin, J., & Beck, Y. M. (1968). *Ward behavior inventory*. New York: Springer.
Burdock, E., & Zubin, J. (1985). Objective evaluation in psychiatry. *Psychiatric reference and record book* (2nd ed.). New York: Roerig Laboratories, Inc.
Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the sixteen personality factor questionnaire*. Champaign, IL: Institute for Personality and Ability Testing.
Cautela, J. R. (1968). Behavior therapy and the need for behavior assessment. *Psychotherapy: Theory, Research and Practice, 5,* 175–179.
Cautela, J. R. (1973, September). *A behavioral coding system*. Presidential address to the seventh annual meeting of the Association for Advancement of Behavior Therapy, Miami.
Cone, J. D. (1977). The relevance of reliability and validity for behavioral assessment. *Behavioral Therapy. 8,* 411–426.
Cone, J. D. (1988). Psychometric considerations and the multiple models of behavioral assessment. In A. S. Bellack & M. Hersen (Eds.), Behavioral assessment: A practical handbook (3rd ed.). Elmsford, NY: Pergamon Press.
Cordes, C. (1983). Mullane: Tests are grounded. *APA Monitor, 14,* 24.
Cronbach, L. J. (1949). *Essentials of psychological testing*. New York: Harper & Brothers (2nd ed., 1960).
Cronbach, L. J. (1960). *Essentials of psychological testing* (2nd ed.). New York: Harper & Brothers.
Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A., (1987). *CVLT: California verbal learning test: Research Edition,* manual. San Antonio: The Psychological Corporation Harcourt Brace Jovanovich.
Edelbrock, C. (1984). Diagnostic issues. In T. H. Ollendick & M. Hersen (Eds.), *Child behavioral assessment: Principles and procedures* (pp. 30–37). Elmsford, NY: Pergamon Press.
Eisler, R. M., & Polak, P. R. (1971). Social stress and psychiatric disorder. *Journal of Nervous and Mental Disease, 153,* 227–233.
Feighner, J., Robins, E., Guze, S., Woodruff, R., Winokur, G., & Munoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry, 26,* 57–63.
Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198.
Foster, S. L., Bell-Dolan, D. J., & Burge, D. A. (1988). Behavioral observation. In A. S. Bellack & M. Hersen (Eds.). *Behavioral assessment: A practical handbook*. New York: Pergamon Press.
Fowler, R. D. (1979). The automated MMPI. In C. S. Newmark (Ed.), *MMPI clinical and research trends*. New York: Praeger.
Gilberstadt, H., & Duker, J. (1965). *A handbook for clinical and actuarial MMPI interpretation*. Philadelphia: Saunders.
Golden, C. J., Hammeke, T. A., & Purisch, A. D. (1980). *The Luria–Nebraska Battery manual*. Los Angeles, CA: Western Psychological Services.
Golden, G. (1981). The Luria–Nebraska children's battery: Theory and formulation. In G. W. Hynd & J. E. Obrzut (Eds.), *Neuropsychological assessment and the school-aged child: Issues and procedures*. New York: Grune & Stratton.
Golden, C. J., Purisch, A. D., & Hammeke, T. A. (1985). *Luria–Nebraska neuropsychological battery: Forms I and II,* manual. Los Angeles: Western Psychological Services.
Goldstein, G. (1979). Methodological and theoretical issues in neuropsychological assessment. *Journal of Behavioral Assessment, 1,* 23–41.
Guilford, J. P., & Zimmerman, W. (1949). *Guilford-Zimmerman Temperament survey*. Los Angeles, CA: Western Psychological Services.
Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry, 23,* 56–62.
Harris, S. L., & Ferrari, M. (1983). Developmental factors in child behavior therapy. *Behavior Therapy, 14,* 54–72.
Hayes-Roth, F., Longabaugh, R., & Ryback, R. (1972). The problem-oriented medical record and psychiatry. *British Journal of Psychiatry, 121,* 27–34.
Haynes, S. N. (1978). *Principles of behavioral assessment*. New York: Gardner Press.
Helzer, J., Robins, L., Croughan, J., & Welner, A. (1981). Renard Diagnostic Interview. *Archives of General Psychiatry, 38,* 393–398.
Hersen, M. (1973). Self-assessment and fear. *Behavior Therapy, 4,* 241–257.

Hersen, M. (1976). Historical perspectives in behavioral assessment. In M. Hersen & A. S. Bellack (Eds.), *Behavioral assessment designs: Strategies for studying behavior change*. Elmsford, NY: Pergamon Press.
Hersen, M. (Ed.). (1988). Behavioral assessment and psychiatric diagnosis. *Behavioral Assessment, 10,* 107–121.
Hersen, M., & Barlow, D. H. (1976). *Single-case experimental designs: Strategies for studying behavior change*. Elmsford, NY: Pergamon Press.
Hersen, M., & Bellack, A. S. (Eds.) (1976). *Behavioral assessment: A practical handbook*. Elmsford, NY: Pergamon Press.
Hersen, M., & Bellack, A. S. (1981). *Behavioral assessment: A practical handbook* (2nd ed.). Elmsford, NY: Pergamon Press.
Hersen, M., & Bellack, A. S. (1988). DSM-III and behavioral assessment. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (3rd ed.). Elmsford, NY: Pergamon Press.
Hersen, M., & Last, C. G. (1989). Psychiatric diagnosis and behavioral assessment in children. In C. G. Last & M. Hersen (Eds.), *Handbook of child psychiatric diagnosis*. New York: Wiley.
Hersen, M., & Turner, S. M. (1984). DSM-III and behavior therapy. In S. M. Turner & M. Hersen (Eds.), *Adult psychopathology: A behavioral perspective*. New York: Wiley.
Hines, F. R., & Williams, R. B. (1975). Dimensional diagnosis and the medical students' grasp of psychiatry. *Archives of General Psychiatry, 32,* 525–528.
Holtzman, W. H. (1958). *The Holtzman inkblot technique*. New York: Psychological Corporation.
Honigfeld, G., & Klett, C. (1965). The Nurse's Observation Scale for Inpatient Evaluation (NOSIE): A new scale for measuring improvement in schizophrenia. *Journal of Clinical Psychology, 21,* 65–71.
Horton, A. M. (1988). Use of neuropsychological testing in determining effectiveness of ritalin therapy in an DDRT patient. *The Behavior Therapist, 11,* 114–118.
Jensen, A. R. (1983, August). *Nature of the white-black differences on various psychometric tests*. Invited address, American Psychological Association Convention, Anaheim, CA.
Kanfer, F. H., & Grimm, L. G. (1977). Behavior analysis: Selecting target behaviors in the interview. *Behavior Modification, 1,* 7–28.
Kanfer, F. H., & Saslow, G. (1969). Behavioral diagnosis. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status*. New York: McGraw-Hill.
Katz, R. C., & Woolley, F. R. (1975). Improving patients' records through problem orientation. *Behavior Therapy, 6,* 119–124.
Kazdin, A. E. (1983). Psychiatric diagnosis, dimensions of dysfunction, and child behavior therapy. *Behavior Therapy, 14,* 73–99.
Klonoff, H., & Cox, B. (1975). A problem-oriented system approach to analysis of treatment outcome. *American Journal of Psychiatry, 132,* 841–846.
Lazarus, A. A. (1973). Multimodal behavior therapy: Treating the "basic id." *Journal of Nervous and Mental Disease, 156,* 404–411.
Lezak, M. (1976). *Neuropsychological Assessment*. New York: Oxford University Press.
Longabaugh, R., Fowler, D. R., Stout, R., & Kriebel, G. (1983). Validation of a problem-focused nomenclature. *Archives of General Psychiatry, 40,* 453–461.
Longabaugh, R., Stout, R., Kriebel, G. M., McCullough, L., & Bishop, D. (1986). DSM-III and clinically identified problems as a guide to treatment. *Archives of General Psychiatry, 43,* 1097–1103.
Machover, K. (1949). *Personality projection in the drawing of the human figure: A method of personality investigation*. Springfield, IL: Charles Thomas.
Marks, P. A., Seeman, W., & Haller, D. L. (1974). *The actuarial use of the MMPI with adolescents and adults*. Baltimore: Williams & Wilkins.
McFie, J. (1975). *Assessment of organic intellectual impairment*. London: Academic Press.
McLean, P. D., & Miles, J. E. (1974). Evaluation and the problem-oriented record in psychiatry. *Archives of General Psychiatry, 31,* 622–625.
Meehl, P. E. (1954). *Clinical vs. statistical prediction*. Minneapolis: University of Minnesota Press.
Menninger, K. a. (1952). *A manual for psychiatric case study*. New York: Grune & Stratton.
Michelson, L. (1984). The role of individual differences, response profiles, and treatment consonance in anxiety disorders. *Journal of Behavioral Assessment, 6,* 349–367.
Michelson, L. (1986). Treatment consonance and response profiles in agoraphobia: Behavioral and physiological treatments. *Behaviour Research and Therapy, 24,* 263–275.
Millon, T. (1982). *Millon Clinical Multiaxial Inventory* (3rd ed.). Minneapolis, MN: National Computer Systems.
Millon, T. (1985). The MCMI provides a good assessment of DSM-III disorders: The MCMI-II will

prove even better. *Journal of Personality Assessment, 49,* 379–391.

Murray, H. A. (1938). *Explorations in personality.* New York: Oxford University Press.

Nathan, P. E. (1981). Symptomatic diagnosis and behavioral assessment: A synthesis. In D. H. Barlow (Ed.), *Behavioral assessment of adult disorders.* New York: Guilford Press.

Nathan, P. E., Zare, N. C., Simpson, H. F., & Ardberg, M. M. (1969). A systems analytic model of diagnosis: I. The diagnostic validity of abnormal psychomotor behavior. *Journal of Clinical Psychology, 25,* 3–9.

Nelson, R. O. (1980). The use of intelligence tests within behavioral assessment. *Behavioral Assessment, 2,* 417–423.

Nelson, R. O. (1987). DSM-III and behavioral assessment. In C.G. Last & M. Hersen (Eds.), *Issues in diagnostic research.* New York: Plenum Press.

Nelson, R. O., & Hayes, S. C. (1979). Some current dimensions on behavioral assessment. *Behavioral Assessment, 1,* 1–16.

Overall, J. E., & Gorham,, J. R. (1962). The brief psychiatric rating scale. *Psychological Reports, 10,* 799–812.

Rapaport, D. (1945). *Diagnostic psychological testing.* Chicago: Year Book Publications.

Raskin, A. (1982). Assessment of psychopathology by the nurse or psychiatric aide. In E. I. Burdock, A. Sudilovsky, & S. Gershon (Eds.), *The behavior of psychiatric patients: Quantitative techniques for evaluation.* New York: Marcel Dekker.

Reitan, R. M., & Davison, L. A. (1974). Clinical neuropsychology: Current status and applications. Washington, DC: V. H. Winston and Sons.

Reitan, R. M., & Wolfson, D. (1985). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation.* Tucson, AZ: Neuropsychology Press.

Robins, S. L., Helzer, J., Croughan, N. A., & Ratcliff, K. (1981). National Institute of Mental Health Diagnostic Interview Schedule. *Archives of General Psychiatry, 38,* 381–389.

Rohde, A. R. (1957). *The sentence completion method.* New York: Ronald Press.

Rosen, A. J., Sussman, S., Mueser, K. T., Lyons, J. S., & Davis, J. M. (1981). Behavioral assessment of psychiatric inpatients and normal controls across different environmental contexts. *Journal of Behavioral Assessment, 3,* 25–36.

Sandifer, M. G., Jr., Pettus, C., & Quade, D. (1964). A study of psychiatric diagnosis. *Journal of Nervous and Mental Disease, 139,* 350–356.

Scales, E. J., & Johnson, M. S. (1975). A psychiatric POMR for use by a multidisciplinary team. *Hospital and Community Psychiatry, 26,* 371–373.

Shneidman, E. S. (1952). *Make a picture test.* New York: Psychological Corporation.

Spitzer, R. L., & Endicott, J. (1977). *Schedule for affective disorders and schizophrenia.* Technical Report. New York: New York State Psychiatric Institute, Biometrics Research Department.

Spitzer, R. L., Endicott, J., & Robins, E. (1977). *Research diagnostic criteria (RDC) for a selected group of functional disorders.* Bethesda, MD: National Institute of Mental Health.

Spitzer, R. L., & Williams, J. B. W. (1983). *Instruction manual for the structured clinical interview for DSM-III (SCID).* New York: New York State Psychiatric Institute, Biometrics Research Department.

Swensen, C. H. (1957). Empirical evaluations of human figure drawings, 1957–1966. *Psychological Bulletin, 54,* 431–466.

Swensen, C. H. (1968). Empirical evaluations of human figure drawings. *Psychological Bulletin, 20,* 20–44.

Szondi, L. (1952). *Experimental diagnostics of drives.* New York: Grune & Stratton.

Taylor, C. B. (1983). DSM-III and behavioral assessment. *Behavioral Assessment, 5,* 5–14.

Tryon, W. W. (1986). Motor activity measurements and DSM-III. In M. Hersen, R. M. Eisler, P. M. Miller (Eds.), *Progress in behavior modification: Vol. 20.* New York: Academic Press.

Tryon, W. W. (1989). Behavioral assessment and psychiatric diagnosis. In M. Hersen (Ed.), *Innovations in child behavior therapy.* New York: Springer.

VanLennep, D. J. (1951). The four-picture test. In H. H. Anderson & G. L. Anderson (Eds.), *An introduction to projective techniques.* New York: Prentice-Hall.

Wechsler, D. (1944). *The measurement of adult intelligence.* Baltimore, MD: Williams & Wilkins.

Wechsler, D. (1987). *Wechsler Memory Scale-Revised.* New York: Psychological Corporation.

Weed, L. L. (1964). Medical records, patient care, and medical education. *Irish Journal of Medical Sciences, 6,* 271–282.

Weed, L. L. (1968). Medical records that guide and teach. *New England Journal of Medicine, 278,* 593–600.

Weed, L. L. (1969). *Medical records, medical education, and patient care.* Cleveland, OH: Case Western Reserve University Press.

White, D. K., Turner, L. B., & Turkat, I. D. (1983). The etiology of behavior: Survey data on behavior therapists' contributions. *Behavior Therapist, 6,* 59–60.

Wolpe, J. (1977). Inadequate behavior analysis: The Achilles' heel of outcome research in behavior therapy. *Journal of Behavior Therapy and Experimental Psychiatry, 8,* 1–3.

Wolpe, J. (1986). The positive diagnosis of neurotic depression as an etiological category. *Comprehensive Psychiatry, 27,* 449–460.

Wolpe, J., & Wright, R. (1988). The neglect of data gathering instruments in behavior therapy practice. *Journal of Behavior Therapy and Experimental Psychiatry, 19,* 5–9.

Zubin, J. (1967). Classification of the behavior disorders. *Annual Review of Psychology, 18,* 373–406. Palo Alto, CA: Annual Review, Inc.

Zubin, J. (1984). Inkblots do not a test make. *Contemporary Psychology, 29,* 153–154.

Winter, D. G. (1987a). Enhancement of an enemy's power motivation as a dynamic of conflict escalation. *Journal of Personality and Social Psychology, 52,* 41–46.

Winter, D. G. (1987b). Leader appeal, leader performance, and the motive profiles of leaders and followers: A study of American presidents and elections. *Journal of Personality and Social Psychology, 52,* 196–202.

Winter, D. G. (1988, July/August). What makes Jesse run? *Psychology Today,* pp. 20, 22, 24.

Winter, D. G., & Carlson, D. G. (1988). Using motive scores in the psychobiographical study of an individual: The case of Richard Nixon. *Journal of Personality, 56,* 75–103.

Winter, D. G., & Stewart, A. J. (1977). Content analysis as a technique for assessing political leaders. In M. G. Hermann (Ed.), *The psychological examination of political leaders* (pp. 28–61). New York: Free Press.

Woods, F. A. (1906). *Mental and moral heredity in royalty: A statistical study in history and psychology.* New York: Holt.

Woods, F. A. (1913). *The influence of monarchs.* New York: Macmillan.

Zullow, H. M., Oettingen, G., Peterson, C., & Seligman, M. E. P. (1988). Pessimistic explanatory style in the historical record: CAVing LBJ, presidential candidates, and East versus West Berlin. *American Psychologist, 43,* 673–682.

Chapter 26

Personality Assessment: A Conceptual Survey

Leonard G. Rorer
Miami University, Ohio

I know nothing stays the same
But if you're willing to play the game
It will be coming around again.
CARLY SIMON

The problem with personality assessment is that there is little agreement as to what one is assessing, much less how to assess it; from this it follows that there is little agreement concerning the domain that it covers. I could start by giving a dictionary definition of personality:

> An individual's enduring persistent response patterns across a variety of situations. It is comprised of relatively stable patterns of action often referred to as TRAITS, dispositional tendencies, motivations, attitudes and beliefs which are combined into a more or less integrated self-structure. (Harré & Lamb, 1983, p. 460)

But we should remember that all such definitions are ultimately circular. All words in the dictionary are defined in terms of other words in the dictionary. We cannot learn the meaning of a word by itself, but only in context. Even dictionaries define words by giving examples of their use. Concepts, including "personality" and "assessment," cannot be defined explicitly; they must be defined implicitly.

I am going to start by assuming that readers have a general understanding of the everyday meaning of these words. There is not much disagreement about that. The disagreement comes when we try to explicate this commonsense understanding and analyze just what we mean by "personality" and "assessment." I take the goal of personality assessment to be the description of people, but only that part of the description having to do with personality. It does not relate to physical appearance or physiological functioning, or behavior as such (a description of how someone ate his or her spinach would not be a personality description, although we might infer personality characteristics from that description); rather, it relates to a person's manner of behaving, his or her moods, and the situations and behaviors he or she chooses as opposed to the ones he or she avoids. This illustrates how our intuitive understanding guides us, and how it is impossible to be explicit about that understanding.

This chapter defines personality assessment, in part, by describing things it includes and things it excludes. As a start, personality assessment applies to such things as intellectual (but not athletic) abilities, aptitudes, attitudes,

interests, and behavioral or personality traits. This chapter also defines personality assessment by describing what personality assessors do, and the rationale that they provide for what they do. However, this is not a how-to-do-it chapter on personality assessment. Readers who want more applied or more detailed instructions could start with Wiggins (1973), an outstanding summary at the time it was published, or the more recent handbook by Goldstein and Hersen (1984). Rather, I try to describe what personality assessors do in the conceptual sense, and to make explicit the often implicit philosophical basis of what has been called personality assessment for the last 60 years or so. I do this historically, but I do not describe particular instruments or techniques. A number of excellent accounts of this kind already exist (e.g., Goldberg, 1971; McReynolds, 1986; McReynolds & Ludwig, 1987).

In order to develop my theme that all personality assessment assumes, and can only be understood in the context of, some model of both personality and assessment, I start with a brief review of some early antecedents. I then describe what I call the "traditional" assessment model. This description includes a fairly detailed look at various strategies of scale construction. Following this, I first review the philosophy of science in the context of which the traditional model was developed, and next summarize some of the modern challenges to that philosophy. Following that, I describe the behavioral and cognitive challenges to the traditional model, more recent developments within the traditional framework, and alternative models that have been developed in the context of the more modern philosophies of science.

EARLY HISTORY OF PERSONALITY ASSESSMENT

Ancient Times to the Early 20th Century

It seems likely that people must have been interested in describing other people from earliest times, and that those descriptions would have included personality as well as physical characteristics and physical (athletic) abilities. The writings of the Greeks from the 5th century B.C. contain personality descriptions based on their humoral theory, which posited four characteristics (sanguinity, irritability, melancholy, and placidity) in terms of which people could be described. These characteristics were thought to be related to the amounts of the four humors in the blood (McReynolds, 1986).

Given the interest that people have in describing one another, it is striking that the development of a systematic means of description is so recent. Early descriptions were tied to beliefs about personality determinants, and were couched in terms of external indicators (astrology), physical descriptions (e.g., a lion-like appearance indicated courage), and hypothesized physiological determinants (e.g., the four humors). The idea of behavioral indicators does not seem to have occurred until the very end of the 17th century, and then only in scattered references (McReynolds, 1986). It was not until the last half of the 19th century that Galton started collecting systematic measurements of behavioral as well as physical characteristics of people: "Galton largely created the bases for the quantitative assessment of individual differences. He pioneered the use of questionnaires, rating scales, word association techniques, test batteries, and other evaluation methods" (McReynolds, 1986, p. 52). He seems also to have been the first to try to find measures of intellectual capacity; he focused mostly on sensorimotor tasks such as reaction time, but also developed the digit span test, a form of which is included in leading intelligence measures to this day.

There were scattered instances of the development and use of checklists and rating scales early in this century. Checklists suggest that there is something that one can have or not have, such as a broken leg or a rash all over the body, or a headache or hallucinations. Rating scales suggest that things exist in some degree or quantity, such as height and weight, or the severity of the headache or the frequency of the hallucinations. Checklists can also apply to things that exist in various degrees, in which case they are really dichotomous rating scales with an arbitrarily selected cutting point. When we use physical descriptors such as "tall" or "short," or psychological descriptors such as "cheerful" or "sad," we are using ascription rules (Rorer & Widiger, 1983) to translate an implicit rating scale into a checklist. An ascription rule tells us how high or low the rating must be on a descriptor before we ascribe it to the person. When we make such an ascription, we say that we have described the person, although we should more accurately say that we have described that aspect of the person (i.e., that we have described the person's height or the person's cheerfulness).

Binet and Woodworth

The next important step forward was taken by Binet and Woodworth, who, although they may not have been the first to have assembled a set of items into a scale, are clearly the ones who popularized the idea. This seemingly simple step represented the start of psychological assessment as we know it today. In the case of the Binet intelligence test, the items were empirically ordered in terms of increasing difficulty to approximate what we would today call a Guttmann scale. The score was taken to represent, at the first level of inference, a child's level of achievement; in relation to the child's age it was taken to indicate the need for remedial education; and, given the additional assumption that children of a given age have had roughly the same educational background, it was inferred to be a measure of the child's intellectual ability or intelligence. If two children who have been exposed to the same educational experiences perform differently, then it is reasonable to assume that the difference in performance is due to differences in ability.

"Ability" is an ambiguous word; so is "ambiguous." A term can be ambiguous either because it is not clear, or because it has more than one meaning and it is not clear which meaning is intended. "Ability" is ambiguous in the second sense. We use it to refer to both a present level of performance ("He can [is able to] run a mile in 8 minutes") and to a potential level of performance ("With the correct training program, he could [has the ability to] run a mile in 6 minutes"). Sometimes the context makes clear which meaning is intended, but in many cases it does not. The picture is muddied further by the fact that ability in the second sense sometimes refers simply to the *fact* that one could learn to do something and sometimes to the *rate* at which one could learn to do it. The controversies having to do with ability testing in general and intelligence (intellectual ability) testing in particular have to do with ability in the second sense.

For our purposes, it is important to note only that the question as to whether a particular test is an achievement test (an ability test in the first sense) or an intelligence test (an ability test in the second sense) is misguided. The same test may be either one, depending on the level of the inference we make. We can say that the individual is able to speak French or to repeat five digits backwards now; that the individual could learn to speak French or to repeat five digits backwards with proper training; or that the person could learn to speak French or to repeat five digits backwards in some (long or short) period of time. The first statement refers to the performance on the test; the second and third are inferences that might be made on the basis of present test performance and other information.

The 116 items in Woodworth's Personal Data Sheet (PDS) constituted a scale of Psychoneurotic Tendencies, which was designed to be used to identify emotionally unstable soldiers in World War I. According to McReynolds (1986), this was the first instance in which a set of items was combined into a single quantitative scale to measure an aspect of personality. The item pool was developed rationally, but the initial set of 200 items was empirically reduced to 116 by eliminating items that were endorsed by more than 25% of normal and inductee samples. This last step was, again according to McReynolds (1986), the first instance in which empirical procedures were used to select items. The elaborate theories of personality inventory construction have all been developed in the little more than 70 years since the introduction of the PDS.

Two significant developments occurred during the 1920s: the development of (1) projective instruments and (2) empirical item selection strategies using an external criterion. As we consider in more detail later, projective instruments used vague stimuli and a free-response format to gather a sample of verbal behavior from which inferences concerning internal psychodynamic structure could be made. Empirically keyed scales were initially constructed on the basis of occupational groups, but starting in the 1930s interest shifted to psychiatrically diagnosed or otherwise identified maladjusted groups; this shift culminated in the publication of the Minnesota Multiphasic Personality Inventory (MMPI) in 1943.

THE TRADITIONAL PERSONALITY ASSESSMENT MODEL

From the present vantage point it seems that psychologists were in much greater agreement as to what they were about 30–50 years ago

than they are today. Psychologists—clinical psychologists, at least—seemed to know what personality assessment was. It is still the case, in the clinical psychology training program of which I am a member, that when a student says that he or she wants to learn to "do assessments," the reference is to the traditional kind that were performed 30–50 years ago, without much thought as to whether there might be other kinds. I am going to describe this model of personality assessment in the context of the 1940s and 1950s, even though I realize that it is still widely held today. What I am referring to as the "traditional" assessment included an anamnesis, sometimes based on a standard form but often obtained by interviewing the client; at least one projective test, usually a Thematic Apperception Test (TAT) or a Rorschach; an objective test of psychopathology, usually the MMPI; some measure of intellectual functioning, probably the Wechsler Adult Intelligence Scale (WAIS) for an adult or the Wechsler Intelligence Scale for Children (WISC) or Stanford–Binet for a child; and an interview focused on the client's problems and presenting complaints. Other tests were often added, depending on the results of these first tests and the questions to be answered. The goal of this assessment was twofold: first, to arrive at a diagnosis; and second, to describe the person's personality structure or "psychodynamics." It was assumed that the diagnosis and the personality structure existed, and that they could be discovered (revealed, understood) by means of the assessment battery.

In the context of the times, the approach made sense. The one clear triumph in understanding mental disorders had been the discovery of syphilis as the cause of general paresis and tabes dorsalis early in this century, and it provided a model for further research focused on identifying disease entities for which treatments could then be sought. Kraepelin (1907/1981) is usually credited with being the first to attempt a comprehensive nosology of psychopathology, and Bleuler (1911/1950) provided an exemplary model in his attempt to describe the schizophrenias. That enterprise continues today with the continuing revisions of the *Diagnostic and Statistical Manual of Mental Disorders* (the latest of which is DSM-III-R; American Psychiatric Association, 1987). This model, which I refer to as the "diagnostic" model, had been successful in medicine for some 50 years. Under this model it was assumed that if diagnostic categories could be established, then appropriate treatments, which would vary from category to category, could be identified.

Although psychodynamic interpretations might superficially seem different, it is important to remember that they were based on the same model. One could *be* a hysteric or an obsessive–compulsive (or, more accurately, one could be a person who had hysteria or obsessive–compulsive neurosis), in the same sense that one could be a consumptive or a person who had diabetes. Certain early life experiences were thought to determine certain personality structures and dispositions, which in turn were thought to determine one's life pattern. There was a limited number of these crucial experiences, and therefore a small number of diagnostic categories. Psychodynamic formulations and the diagnostic system were inextricably intertwined. Given a diagnosis, one could infer a developmental personality pattern, and vice versa. Given the assumption of psychological etiology, the purpose of the psychodynamic formulation was to identify the kinds of life experiences on which psychotherapeutic treatment should focus. In conjunction with the life history, the psychodynamic formulation might target the particular life experiences that were crucial for this client, or might indicate that the particular experiences remained to be discovered.

The traditional assessment procedures are all tied to the diagnostic model. Meehl and Golden (1982) have provided a clear exposition of the logic by which diagnostic categories are established and the methods by which individuals are assigned to such categories. The recent history of acquired immune deficiency syndrome (AIDS) provides a familiar example of how nosological categories are identified. It was first noticed that there had been an unexplained increase in the incidence of Kaposi sarcoma; next that these cases tended to be confined to a particular category of the population (gay males); and then that gay males seemed to be developing a rather vague disorder that might be described as general malaise (fatigue, fever, etc.), and that they also seemed to have a decreased resistance to other infections. When this was combined with reports of a similar disorder in Africa, it seemed that there was a sufficient reason to think that medicine was dealing with a genuine disease entity, and a search was launched for an etiological agent (in this case the presumption was high that it was a virus). Most believe that the virus has now been identified, but there are still some skeptics who believe that the disorder is caused by a pair of viruses (only the less lethal of which has been identified), and others who believe that it is not caused by a virus at all.

There are at least two tests to identify antibodies to the virus, but they are imperfect, and one may have the virus for some period of time (no one knows how long) before one develops antibodies. The diagnosis is based on the tests, the person's history, and the other symptoms that the person has, and is therefore probabilistic. It is known that the virus can be transmitted through the blood, and it is believed that it can be transmitted by sexual contact (though the mechanism of transmission is not known, and some individuals have had hundreds of sexual contacts with persons carrying the virus without themselves getting it). It is not known whether the virus can be transmitted in other ways. Thus, we are still learning about the disease; as we learn, the methods of diagnosis and treatment, as well as the definition and understanding of the disorder, change. It is an open concept.

Once an etiological agent has been identified, then a positive diagnosis can be made by associating the agent with the person. Prior to determining the cause of a disorder (and to some extent afterward, as well) we are operating on the basis of a hypothetical construct, an open concept, for which there is no explicit definition and no infallible (pathognomonic) indicator. In the absence of such an indicator, we are forced to make diagnoses on the basis of probabilistic signs and symptoms. It is not possible at this point to specify any combination (conjunction) of signs that must be present, or any set (disjunction) from which a certain number will suffice. (Certain sets or combinations may be arbitrarily specified for research or reporting purposes, as in the case of DSM-III-R). Rather, it is assumed that the more positive indicators one has, the greater the probability that one falls into the category. It may also be the case, or may be assumed to be the case, that the number of positive indicators is related to the severity of the disorder, if it is a disorder that can vary from mild to severe. These are not necessarily unrelated, because the more severe the disorder, the more signs and symptoms it would be expected to produce; the clearer the signs and symptoms would probably be; and therefore the higher the probability with which it could be identified.

The traditional assessment procedures followed from this model. Anamnesis responses, personality inventory items, and projective test scores were treated as diagnostic signs. The number of positive responses was summed, and this number was assumed to be related to the probability and the severity of the disorder to which the signs were presumed to be related. Within the context of the prevailing psychodynamic model, the psychodynamic formulation was intended to identify the specific etiological agents associated with the diagnosis in this individual's case.

Projective Tests

Projective tests were characterized primarily by the use of ambiguous stimuli and free-response formats. To the extent that they produced quantitative scores, the logic of their use was no different from that of personality inventories. The Rorschach, for example, could be scored for response categories such as Form, Color, and Human Movement. These scores were initially related to diagnostic categories on theoretical grounds, but the relations were not supported by subsequent empirical studies.

In addition, the free-response protocols could be interpreted in terms of the theory, which assumed that one's basic personality structure determines the nature of one's responses in all aspects of one's life. It followed that if a tester used vague or ambiguous stimuli to which a person could respond in a variety of ways, then the particular response that was selected must be a function of the person's dynamic structure. Stimuli that were suggestive of basic response categories but still vague enough to be interpreted in several different ways were thought to provide the best opportunity to trigger the crucial, revealing responses.

The validity of the process obviously depends crucially on the verisimilitude or utility (depending on one's epistemological position) of the theory. It also depends on the accuracy of the subjective scoring of the protocols for psychodynamic content—a process requiring that the response be interpreted in terms of the theory. This creates an infinite regress, because the responses are now, for the interpreter, a form of projective test, and there is no way of knowing whether the interpretation is a func-

tion of the psychodynamics of the respondent or the interpreter (see Meehl, 1982, for an analysis of this problem with respect to psychoanalytic interpretations in general). Evidence of high agreement among interpreters would help deal with this problem, though it would not solve it. In any event, the inter-interpreter reliabilities are not high, so the problem remains.

In my view, the huge literature on the psychodynamic interpretation of these instruments has shown their utility to be negative. Of more relevance to this chapter, I do not know of any developments in the assessment theory underlying the subjective interpretation of these instruments in the context of psychoanalytic theory in the last 50 years. Therefore, I have nothing more to say in this chapter about projective techniques as such, and move on to the discussion of personality inventories. I return to the problem of the interpretation of protocols in the sections on cognitive assessment and narrative analysis.

Personality Inventories

By "personality inventory" I mean a set of stimuli (usually verbal items) to which the respondent is instructed to respond by means of a limited number of structured response options. Various techniques are used to identify sets of items that form scales. Here, I describe the underlying logic of seven of these scale construction strategies.

The Rational Strategy

In general, if one wants information from someone, the easiest and best way to get it is to ask directly. As we have seen, the Woodworth PDS was constructed on this basis. Its goal was to identify inductees who might be unfit for military service, and it was therefore composed of items describing psychological conditions that might render someone unfit for military service. Such direct items are the most efficient means of obtaining information if the respondent (1) can understand the question, (2) knows the answer to the question, and (3) is willing to answer the question honestly. In the case of the PDS, it was obviously easy for a respondent to malinger, and there was an equally obvious motivation for many of them to do so.

The PDS took a step forward by assembling a set of such items into a scale. The implicit assumptions were that the items were all related to the same construct (psychological fitness for military service), and that the more such items one endorsed, the less likely it was that one was fit for service—in other words, that there was a monotonic relation between scores on the scale and fitness for military service.

The designation of this means of scale construction as "rational" may be something of a misnomer. In the modern sense of this word, "rational" refers to thoughts and actions that are reasonable in the light of well-established empirical knowledge. Its use here is in reference to the historical distinction between rationalist and empiricist philosophies. In that use it was meant to contrast rational with empirical approaches to scale construction. The contrast is one of disapprobation, and the emphasis is on the fact that rational scales were constructed without empirical checks on their validity. Thus, "rational" here refers to acting in the absence of, or with disregard for, empirical information.

The Rational–Theoretical Strategy

One may rationally select items for a scale, not because one is interested in the answer to the item itself, but because the response is thought to be related to a construct in which one is interested. If, for example, one is interested in measuring human needs (e.g., Murray, 1938), then one may include an item of the form "Whenever I have a new sheet of stamps, I always carefully remove the border from an outside stamp before using it"—not because one really cares whether the person uses stamps with the border on them, but because one thinks the response might be related to a construct called Need for Order.

There is obviously no clear line between rational and rational–theoretical scales. Each of the items in the PDS was thought to be significant in its own right, but they were also thought to be related to the construct of Psychoneurotic Tendencies. However, the latter construct was not of theoretical interest in the same sense as something like Need for Order, which is embedded in a theory. The distinction really refers to the investigator's orientation and purpose: Is one interested in the answer itself, or in the answer as an indicator of a construct from a particular theoretical orientation?

Rational and rational–theoretical means of scale construction fell out of favor from the 1930s through the 1960s. Beginning in the 1970s, however, the cycle that had reached the empiricist extreme during the 1940s and 1950s returned to the point at which rational–theoretical approaches were described as the preferred first step in personality inventory scale construction, within the context of a sophisticated construct validation model (Jackson, 1971).

The Internal-Consistency Strategy

In the absence of an external criterion that the scale is intended to predict, are there ways in which one might improve on the procedures described above? If one's theory is that the construct being measured is a unitary construct, then one may believe that all items in the scale should be positively correlated with one another. Internal-consistency coefficients (e.g., coefficient α) or correlations between each item and the total of the remaining items may be used to insure scale homogeneity. In this case, empirical information concerning the relations among responses is being used as one kind of check on the adequacy of the items to measure the intended construct. However, nothing in the procedure guarantees that the items are measuring the intended construct, as opposed to some other construct. Nor is there any guarantee that they are measuring a single construct; they may be measuring two or more constructs that happen to be positively correlated in the sample being studied. In fact, it is possible to get high α coefficients even for a scale measuring two constructs that are uncorrelated in the population (O. P. John, personal communication, August 9, 1989).

Investigators who adopt the internal-consistency approach believe that a scale should measure only one construct, and that this construct should be homogeneous. If the construct can be broken into constituent parts, then a scale should be constructed to measure each of those parts. The scales can then be combined either statistically or subjectively to provide a measure of the superordinate construct.

Although a great deal has been written about the desirability of scale homogeneity, the limits of this desirability are only rarely considered. If all items correlated perfectly, then obviously one would need only one item, not the whole scale. If the items correlated at the limits imposed by their individual reliabilities, then additional items might add reliability, but no new information. If one combines items in a scale, one is implicitly acknowledging that no individual item is designed to measure the construct as such, but rather only some aspect or instance of the construct. Items measuring different aspects or instances of a construct obviously should be less than perfectly correlated after one allows for less than perfect reliability. How much less remains an unanswered theoretical question, probably because, as a practical matter, test constructors have not been confronted with the problem of items that seemed too highly correlated.

The Empirical, Criterion-Group Strategy

None of the procedures described above uses empirical data to assess the relation between item responses and the construct of interest. Empirical, criterion-group keying is straightforward: Assemble a pool of items, administer them to the group one wishes to describe (the criterion group), and key them in terms of the predominant response of the group. Items can be differentially weighted to reflect their endorsement frequencies. Individuals who take the scale can then be described in terms of the extent to which their answers correspond to those of the group. Note that if one considers the prototypical member to be the one who has the most characteristics that identify the group, then the prototypical member may receive a score higher than the average for the group (i.e., he or she would or will be more like the group prototype than the average member of the group). Thus, as scores on the scale increase, they first indicate increasing similarity to a typical member, and then decreasing similarity. Scores are linearly related to group prototypicality, but not to group representativeness. (In some groups, such as that of people with schizophrenia, the prototypical member is not one who has every possible schizophrenic sign). Not many scales have been constructed this way.

The Empirical, Contrasted-Groups Strategy

The empirical, contrasted-groups method is the strategy that is most often meant when writers

refer simply to "empirical scale construction." The shorthand designation causes few problems because most empirically keyed scales were developed by the contrasted-groups strategy. In this procedure, an item pool is administered both to a group of interest (the criterion group) and to a group from whom they are to be differentiated (the reference group). The latter is generally a group representative of people in general, but may be a narrower group representative of, say, psychiatric hospital patients, and may in principle be any other group. One retains those items that are answered signficantly differently by the members of the two groups. This procedure should be cross-validated (although it seldom is) to remove those items that may have initially been included because of Type I errors.

The items that are included by these two empirical procedures are obviously very different—a fact that is sometimes overlooked in interpreting responses to scales that have been developed by these empirical procedures. The first method retains items that are common to the two groups; the second does not. For example, in the physical realm the former procedure includes items relating to having two arms, two legs, and two eyes, whereas the latter does not. Depending on the item pool with which one begins, a criterion-group scale may be composed entirely of Barnum-type statements—that is, statements that are true of almost everyone. Thus, the scale may not be very descriptive of the criterion group in the sense that it contains no information to distinguish members of that group from any other group.

The contrasted-groups procedure, on the other hand, may retain items that are not representative of the criterion group, because, even though they are answered by only a minority of the criterion group, that minority is larger than the minority of the reference group who endorse the item. For example, an item such as "I sometimes see things that are not there" may be answered by only a minority of a group of psychiatric patients, but may be answered by almost none of the members of the reference group, and so will be included in the scale, even though it is not representative of the criterion group. The more nearly similar the criterion group is to the reference group, the more misleading the items will be as a description of the criterion group, because only those items relating to the small area of difference will be retained. The criterion group will be caricatured by items that describe only a fringe minority or relatively unimportant characteristics of the group.

Both rational and empirical scales have been developed on the basis of two distinctly different population models. One posits distinct groups, or "taxa" (e.g., Meehl & Golden, 1982). The other posits traits or characteristics that are assumed to be normally distributed in the population. In either case, there must be some independent way in which to identify criterion-group members so as to be able to construct the scale, although once the scale is constructed it may turn out to be a better indicator than the one that was used in the first place (Dawes & Meehl, 1966; Meehl & Golden, 1982).

The way in which the criterion group is constructed may pose a problem for subsequent interpretation. Consider the continuous case first. Here the usual procedure is to identify individuals falling at the extreme ends of the continuum and to select items that discriminate between these two groups. Unfortunately, those items that best discriminate individuals who fall at the extreme ends of the continuum may not discriminate among those who fall toward the center, and it is the latter for whom the scale is most needed, because those at the extremes can probably be identified without use of the scale. The same problem occurs in a slightly different form when distinct groups are assumed to exist. Here, in an effort to get a "clean" criterion group one selects only clear-cut cases of, say, schizophrenia. The resulting scale may well identify such clear-cut cases, but, again, one does not need the scale for that. Rather, one needs the scale for assistance in diagnosing borderline cases, and there is nothing in the procedure to guarantee that the scale will be very good at that.

In contrasts of the rational and the empirical approaches, the attention paid to item content has often been overemphasized. It is true that the rational approaches utilize item content exclusively, and that the ultimate basis of empirical scale construction strategies is endorsement frequency. But it is not true, in general, that content has been ignored in assembling item pools for empirical scale construction. Vocational inventories contain items relating to occupations and interests; personality (or should we call them psychopathology?) inventories, such as the MMPI, contain items relating to psychological problems; and so on. One of the reasons why the MMPI has endured as long as it has is probably because Hathaway had an unusual sensitivity to what his patients said. Other inventories with items that seem more clever or original have not fared as well.

The Empirical, Itemmetric Strategy

The fact that item content has typically been considered in assembling item pools for empirical scales is made clearer by considering those cases in which scales have been constructed on the basis of criteria other than item content. Examples include endorsement frequencies, social desirability scale values, and differential endorsement frequencies. In the first two cases the goal has been to develop scales that would measure acquiescent and social desirability response tendencies. Unfortunately, the interpretation of scales constructed in this fashion turns out to be problematic, because neither endorsement frequency nor social desirability is independent of item content, or of each other. In the third case the goal has been to test the radical empiricist theory of scale construction. I comment briefly on each of these in turn.

Response Styles. Unless one adopts the radical empiricist approach, one assumes that the responses to personality inventory items are responses to the content of those items. The "response style" hypothesis stated, first, that individuals might be responding, not to the item content, but rather on the basis of a preference for some response category, such as "true" or "false"; and, second, that such response preferences might be enduring tendencies reflective of a basic personality style. The hypothesis stimulated a great deal of research aimed at building measures of response styles, particularly acquiescence response style, so as to use these scales to remove the stylistic components from responses and get purer content measures.

The first part of the hypothesis seemed to be supported when scales constructed on the basis of endorsement frequencies were shown to account for sizable proportions of the response variance. Given the experimental assumption that such scales would be composed of diverse content, and would therefore not be subject to content interpretations, it was inferred that these scales must measure response styles. The assumption turned out to be false. The methodologically superior approach of item reversal showed that such scales were more accurately interpreted in terms of their content, and that the stylistic contribution to variance in personality inventories was small. There was never any support for the hypothesis that stylistic tendencies represent enduring personality characteristics (see Rorer, 1965, for a detailed review).

Nonetheless, it has become part of the enduring wisdom that "standard practice" calls for the construction of balanced scales. "Balanced" sometimes seems to mean an equal number of true- and false-keyed items, and other times to mean an equal number of positively and negatively phrased items. It is my impression that there is not a clear appreciation of the fact that these are not the same. In the first case one could have a scale of positively phrased items, half of which were keyed true and half false, whereas in the second case one could have a true-keyed scale composed of an equal number of positively and negatively phrased items. Acquiescence response style, renamed "agreement acquiescence" by Bentler, Jackson, and Messick (1971), would be controlled by the former, but not the latter.

However, Paulhus (in press) notes that "The original proponents of the importance of acquiescence, Messick and Jackson, shifted their focus to acceptance acquiescence." The latter refers to the tendency to ascribe many characteristics (even seemingly contrary ones) to oneself, without respect to keying. The importance of acceptance acquiescence was challenged by Block (1971), and has received little attention. Paulhus (in press) was able to find only two personality instruments designed to measure or control for it, and both of these are measures of socially desirable responding. Thus, balanced keying has come to be part of the accepted wisdom, even though the original proponents "concluded that the effects of agreement acquiescence are insignificant" (Paulhus, in press).

Social Desirability. In contrast to "response styles," which relate to a tendency to prefer a certain response option independently of the item content, "response sets" refer to a tendency to respond to item content with the intent of presenting a certain picture of oneself. When it is used in the literature, "social desirability" usually seems to refer to a response set—a refer-

ence that is made clear by "social desirability response tendency" or some similar designation. However, it should be emphasized that "social desirability" is ambiguous in the sense of having two distinct meanings. One is the one that has just been suggested: a tendency to put oneself in a good light, to give the socially desirable response, even if it may not be the most accurate one. The tendency may be either conscious or unconscious. In either case, it is a form of dissimulation.

One would like to be able to partial these response tendencies out so as to get a better estimate of what a correct content response would be. Dissimulation scales have been constructed in two ways: (1) empirically, by contrasting the responses of people presumed to be making themselves look good or bad with people presumed to be answering honestly, or at least more honestly; and (2) rationally, by selecting items judged to be either desirable but unlikely to be true, or undesirable but likely to be true.

The early work which popularized the notion of social desirability was based on the Social Desirability *(SD)* scale, constructed from the MMPI item pool by selecting items judged to be either socially desirable or undesirable, without regard to their likely veridicality (e.g., Edwards, 1957). Edwards contrasted his scale with the MMPI Lie *(L)* scale, and noted that they loaded on different factors (I and III, respectively). He stated simply that the *SD* scale measured socially desirable responding (a potentially misleading designation) and that responses could be accounted for by the social-desirability scale values (*SDSV*s, determined by having subjects rate the social desirability of the items on a 9-point scale) of the items. In this second sense, social desirability is not a response set; it is simply a psychometric fact that is neutral with respect to the veridicality of the responses.

However, social desirability has not been construed in terms of this narrow psychometric definition. It has been claimed that the first factor of the MMPI measures social desirability, presumably in contrast to content, and that desirability scales should be used to partial social desirability out of other scales—a suggestion that makes sense in relation to social desirability in the first sense as a response tendency, but makes no sense in relation to social desirability in the second sense as a psychometric fact.

Some behaviors and characteristics are more desirable than others. We cannot measure behaviors and characteristics independently of their social desirability. Any form of psychopathology provides an example. One of the reasons why we are interested in measuring anxiety, depression, psychoticism, or any other form of psychological maladjustment is precisely because such maladjustment is not desirable. If we were to partial social desirability out of psychoticism, what we would presumably have left would be psychoticism that is not socially undesirable. What, pray tell, would that be? Certainly not what any clinician is interested in measuring. Attitudes, behaviors, beliefs, characteristics, dispositions, tendencies, and traits do differ in their social desirability, and a measure with that desirability partialled out is simply a measure of something other than the original construct. To remove that part of the meaning that is associated with social desirability is to change the meaning.

In the more than 30 years since Edwards popularized research into social desirability, investigators have first lumped all the dissimulation and desirability measures together under the general rubric of "social desirability," and then rediscovered that there are really two kinds of measures there (e.g., Paulhus, 1986). There is not much disagreement about the first kind, though it has been suggested that terms such as "lie" and "dissimulation" are overly pejorative and should be replaced. Paulhus (1986) has suggested "impression management."

There is continuing disagreement about the interpretation of the second kind of measure. However, no one has improved on Block (1965), who showed empirically that peer ratings indicated that item content describes respondents accurately. And no one has responded to Norman (1967), who pointed out that the method of scale construction precludes inferences concerning response tendencies on logical grounds (see the section on "Normative, Ipsative, and Idiothetic Strategies," below). Paulhus (1986, in press) concluded that the second kind of measure probably does measure self-deception, but only to the extent that such self-deception is part of normal healthy functioning; he noted that social desirability in this sense should not be partialled out of personality measures, because it is part of what the personality scales are designed to measure. Thus, whether or not one believes that the scales include a component of self-deception, the conclusion is the same: There is no justification for using these scales to partial out social desirability.

Radical Empiricism. The radical empiricist position, dubbed the "deviation hypothesis" by Berg (1967), stated that scales composed of items with no obvious content relation to the construct being measured, or scales composed of sets of geometric drawings with no content at all, could be constructed to perform as well as scales with relevant content. The proposed methodology was a straightforward, brute-force application of the empirical, contrasted-groups strategy. All that was necessary was to assemble a sufficiently large pool of potential items and a sufficiently large group of subjects and let the computers do the rest. If such scales could have been constructed, they would have solved one of the long-standing problems of personality assessment, because they would have been virtually immune to dissimulation. Unfortunately, after some apparent early successes had raised hopes, all such attempts ultimately failed (e.g., Goldberg & Slovic, 1967; Norman, 1963). The failure of this ultimate test of the empiricist approach during the 1960s undoubtedly played a role in the increasing emphasis on theoretical and conceptual coherence during the 1970s and 1980s.

The Factor-Analytic Strategy

The logical extension of the principle of scale homogeneity is exploratory factor analysis, especially if one is constructing a set of scales rather than a single scale. In contrast to the rational and empirical strategies, which assume that one knows what one wants to measure, the factor-analytic strategy assumes that one does not know what one wishes to measure. The goal of the procedure is first to determine that, and then to build measures of the constructs identified by the factor analysis. The model assumes that one is identifying dimensions along which people differ, and is therefore not appropriate for distinct groups (taxa).

Conceptually, there are two stages of the factor-analytic process: One first determines how many dimensions (factors) are needed to reproduce the responses to the items, and then how best to position the dimensions (axes) so as to provide the most meaningful or interpretable outcome. The second step requires a decision that is important both conceptually and mathematically, but may not make a lot of practical difference—namely, whether to require the dimensions to be orthogonal or to allow them to be correlated. Orthogonal solutions are mathematically simpler, and allow one to calculate the independent contribution of each factor to each variable. However, such a solution is not desirable if the most "natural" solution involves factors that are, in fact, correlated in the population one is studying. To use another example from the physical realm, if one had a set of variables related to the height and weight of people, and if one wanted a two-factor solution that would describe the relations among these variables with respect to the dimensions of height and weight, then one could not insist on an orthogonal solution, because height and weight are correlated in the general population. In an orthogonal solution, if the first factor represented height, then the second factor would represent that aspect of weight that is independent of height—namely, the tendency of people to be fat or thin for a given height.

Although there are algorithms that will perform both steps on the basis of designated criteria, thereby making the process both objective and empirical, sophisticated practitioners often run numerous analyses using a variety of algorithms to see which gives the "best" solution, or attempt to improve on the results by making subjective adjustments. The primary criteria seem to be the meaningfulness and comprehensibility of the result in terms of previous findings and the investigator's theoretical beliefs. Once the dimensions to be measured have been decided upon, the items that correlate most highly with each dimension can be empirically selected to construct scales to measure those dimensions (see Comrey, 1988, for a very readable description of his factor-analytic reasearch program).

The construct is now defined by the content of the items in the scale. The investigator may, on the basis of that content, argue that the construct corresponds to a construct in some theoretical framework, or he or she may argue that a new construct, requiring a new theoretical formulation, has been discovered. If the result were to be interpreted within the traditional framework that I have described, then one would have to assume that there are certain dimensions that one is trying to identify (discover). Few factor-analytic investigators would interpret their results that way these days; they would probably opt instead for a constructivist approach within which it would be acknowledged that a number of different descriptions

might be possible, with the selection ultimately to be made on the basis of pragmatic criteria.

Test Interpretation

Even if a scale has been empirically constructed by either the criterion-group or the contrasted-groups method, we still do not know, except by looking at the content of the items, what characteristics of the group may be involved in making the distinction. Thus, the means of interpreting or understanding an empirically constructed scale is not a great deal different from that of a rationally or factor-analytically constructed scale. Furthermore, consulting the content of the items does not tell us what other characteristics, not represented in the item pool, may be related to those represented in the scale. Thus, no matter how a scale has been constructed, the only way in which to know what it indicates is to investigate its empirical correlates. The same reasoning applies to sets or patterns of scales.

Empiricist investigators therefore began studying groups of people who scored high or low on a given scale. The characteristics of such a group could turn out to be similar to those of the group on the basis of which the scale was constructed. Indeed, the scale might turn out to be a means of selecting a more homogeneous group (Meehl & Golden, 1982). But the characteristics might turn out to be very different.

In addition, it was observed that some patterns of MMPI scale scores occurred much more often than others. This raised the hope that such patterns might be the road to the discovery of true diagnostic groups; that is, that the groups identified by these patterns might provide a new and improved diagnostic system. The result was a proliferation of "cookbooks" describing a number of these groups in different populations (e.g., Gilberstadt & Duker, 1965; Hathaway & Meehl, 1951; Marks & Seeman, 1963). Although the hope of a new diagnostic system was not realized, the science of test interpretation was fundamentally altered. Related studies made it clear that logically reasoned interpretations were no match for empirically derived ones. And psychologists were forced to confront the problem of Bayesian inverse probabilities: The probability of being schizophrenic, given a high score on a schizophrenia scale, is not the same as the probability of having a high score on the schizophrenia scale, given that one is schizophrenic (e.g., Dawes, 1962; Rorer, Hoffman, Laforge, & Hsieh, 1966).

If one has identified the criterion one wishes to predict, then the logical extension of the finding that the empirical interpretation of scales is superior to subjective interpretation is that the empirical selection of scales may be superior to subjective selection, and it turns out that it is. One should therefore have the computer determine empirically which scales should be included in the prediction set and how they should be weighted. Because the number of possible scale combinations is large, the number of possible empirical correlates endless, and the number of possible populations limited only by the investigator's imagination, modern computerized cookbooks rely on a combination of empirically derived and clinically inferred descriptors. Butcher (1987) provides an up-to-date survey.

It is a historical fact that such cookbooks have been constructed almost entirely on the basis of the MMPI, although there is no logical reason why that should be the case. Is there an empirical one? More generally, if one is building an empirical prediction system, does it matter whether the scales have been assembled rationally, empirically, or factor-analytically? In a pioneering study, Goldberg (1972) asked just that question. The surprising answer, confirmed in subsequent studies (Burisch, 1986), is that it does not. Scales assembled randomly or solely on the basis of itemmetric considerations (*SDSV*s or endorsement frequencies) have little validity, but the other systematic means of scale construction seem about equally effective, at least with cooperative subjects. The results in other settings may well be different. Thus, when the empirical scale construction strategy is carried to its logical conclusion, it has an ironic result: Empirical scale construction results in no predictive advantage; one can do as well assembling scales rationally, and one can do it a lot more cheaply. The same is not true of scale interpretation, where the evidence in favor of empirical interpretation is overwhelming.

LOGICAL EMPIRICISM AND ITS CRITICS

To understand the traditional assessment model, it is necessary to keep in mind some of the tenets of logical empiricism, in the context of which the traditional assessment model was developed. That position has not been accepted in philosophy for at least 25 years, and is finally beginning to lose support in psychology. In this and the following section, I indicate some of the problems with the empiricist position as I describe it, and briefly describe some more modern alternatives.

1. In order to be scientifically acceptable or meaningful, statements had to be "objective," by which it was generally understood that they had to refer to something "observable." The logical empiricists realized early on that such a criterion was much too strict, and readily accepted statements that were only indirectly confirmable. Physicists have never had any reluctance to talk about things that are in principle unobservable (e.g., atoms, quarks, black holes), but psychologists, thinking that they were being scientific, competed with one another to see who could be the most rigorous—that is, the most objective.
2. Given the goal of objectivity, the problem was to make that which is subjective, objective. There is, ultimately, no way to do that. Nevertheless, it was thought that unobservable entities could be dealt with by means of operational definitions. Unfortunately, it has never been possible to give an acceptable formulation of an "operational definition," and most philosophers of science now agree that it cannot be done (Suppe, 1977). Even if it were possible to formulate operational definitions, one has to have some way of determining the extent to which the operational definition corresponds to the subjective experience, and there is no objective way to do that. One solution is to pretend, as some behaviorists did, that one can ignore subjective experience, and that one should put an end to the traditional personality assessment enterprise (see below).
3. It was believed that it was possible to obtain theory-free data, and that assessment procedures could, and therefore should, be theory-free. It was only after the demise of empiricism in philosophy, and the behavioral and cognitive revolutions in psychology, that we were able to appreciate that all assessments and all assessment procedures are meaningful and understandable only in a theoretical context.
4. The goals of scientific research were thought to be prediction and control. Prediction and control required that one know the cause of an event. To establish the cause of an event, and therefore to be able to predict it, one had to show that it was a special case (instantiation) of a general covering law. The task of the scientist was to develop theories (laws) from which deductions could be made, and then to test those predictions experimentally. If the experimental result corresponded to the prediction, then the theory was said to be confirmed, the event was said to be explained, and the law was taken to state the cause of the event. Acceptable scientific research was narrowly limited to laboratory or controlled experimentation. Observational or correlational procedures were rejected, because it was thought that only controlled experiments could establish causal relations.

A realist position allows a far different view of causality. The goal is not to discover general laws, but to describe structures. Causal explanations are provided by a description or account in some theoretical context, and can be of many kinds. Controlled experimentation is not the only, or even the most desirable, means of providing causal explanations. The gene, our understanding of which has changed radically over the years, explains the phenomena of inheritance. The double helix provides an explanation at another level. Both are structures. Astronomical theories explain the movements of the stars and planets, which obviously cannot be studied experimentally in the laboratory. Thus, both the goals and methods of science have changed. (For a more detailed summary of the philosophy of science as it relates to psychology, see Rorer, in press.)

The empiricist view of science created a tension within the field of personality assessment. The instruments developed in the diagnostic context were objectively scorable and quantitative, and therefore scientifically respectable. The projective tests, developed within the psychodynamic framework, were not. Some who worked within the latter framework tried to make their research enterprise conform to the sicentific model; others rejected it, as they should have, but had no strong candidate to put in its place.

CHALLENGES TO THE TRADITIONAL MODEL

As it turned out, the entire assessment enterprise was to be attacked—first from the right by the behaviorists, who claimed that it was not

sufficiently scientific, and then from the left by the cognitivists, who claimed that it was too constricted by an outmoded philosophy of science to be able to deal with the problem to which it was addressed. Here, I first consider the behaviorist attack, which was launched within the empiricist framework.

Behavioral Assessment

Although behavioral methods of treatment enjoyed some popularity during the 1920s as a result of the writings of Watson, behavior therapy or behavior modification did not become a recognized speciality until the 1960s, when it was embraced by academicians as a scientifically respectable alternative to psychoanalysis for the treatment of psychological (behavioral) problems. Hayes (1986) distinguishes among three kinds of behaviorists. "Traditional methodological behaviorists" accept the early logical empiricist notion that science can deal only with that which is publicly observable. "Contemporary methodological behaviorists," in parallel with the liberalization of the logical empiricists' meaning criterion, would include the study of "inferred events that are thought to exist at another level but that cannot be addressed directly by science" (Hayes, 1986, p. 39). "Radical behaviorists," following Skinner, reject public observability as a criterion of scientific meaningfulness, and focus instead on the contingencies surrounding an observation. In contrast to methodological behaviorists, radical behaviorists reject logical empiricism as a philosophical basis. It is a contextualist position, focusing on the subjectively determined act in context (Hayes, 1986).

In my reading of the literature, behavior therapists and behavior assessors seem to invoke methodological rather than radical behaviorism as the justification for their procedures, even though there may be references to Skinner. The general claim is that behavior therapy is objective, because it relies upon overt behavior, whereas psychoanalysis is not (Freud's claim to the contrary notwithstanding), because it focuses upon private events; behavior therapy is therefore claimed to be more scientifically respectable than psychoanalysis or any of the related psychodynamic theories (Kazdin, 1979). This belief was widely shared in academic circles in the 1960s and 1970s, and it has survived, albeit less strongly, through the 1980s.

Behavioral theorists reject internal structures or dispositions as causes of behavior, and therefore reject both internal dynamics and disease processes as accounting for disordered behavior. Because they reject the disease concept, they reject the diagnostic model according to which people could be put in categories. Disordered behavior is not a sign of an internal dysfunction; it *is* the dysfunction. The task of the behavioral assessor, therefore, is not to obtain data on the basis of which inferences can be made to internal states, dispositions, or events, but rather to obtain samples of the behavior to be modified.

Note that this fundamental difference in theoretical approach does not necessarily imply any difference in the data that one may collect, but only a difference in the way the data are treated once they have been collected. Both approaches may use self-reports, behavioral rating scales, objective tests, and physiological reactions—in one case treating them as signs of the disorder to be corrected, and in the other as samples of the behavior to be modified. Nevertheless, the change in focus from covert to overt behavior and from signs to samples resulted in a different emphasis on the kinds of data to be obtained. The behaviorists were interested in observing *in situ,* or in constructing situations in which the problematic behavior would be exhibited, or in obtaining verbal reports of the behavior to be changed.

Thirty years of research on test development and theory had been based on the premise that one could not just take behavior, especially verbal reports, at face value. It was this research on which the claim of scientific respectability for clinical psychology had been based. With the behaviorists' emphasis on behavioral contingencies, this work was seemingly undone at a stroke, again in the name of scientific respectability.

Just how different are these positions? Not as different as it might seem (e.g., Meehl, 1986). First, it should be noted that, in general, behaviorists do not study behavior, but rather the results of behavior (lever presses, maze turns, quantity of food eaten, etc.). Second, behavior as such is not observable. Movement is, but it requires inferences to convert movement to behavior. Third, even though they might deny it, I think many behaviorists do consider their observations, such as heart rate, blushing, perspiring, and trembling, to be indicators of internal states—in this case a state of anxiety.

Cognitive Assessment

"Cognitive assessment" is another ambiguous designation. Until the mid-1970s, at least, it referred either to intellectual assessment, usually with reference to the administration of an IQ test, or to neuropsychological assessment, which was understood to mean the assessment of cognitive or intellectual dysfunction (brain damage) as a result of trauma, illness, or degeneration. That is, it referred to the assessment of either cognitive abilities or the impairment of cognitive abilities. In this sense it may be taken to refer to the assessment of cognitive processes—in particular, to the quality or efficiency of cognitive processes. In the present context, however, "cognitive assessment" has quite a different meaning. It refers to the assessment of beliefs—the contents, both conscious and unconscious, of the mind, including the contents of cognitive (reasoning) processes. It includes the assessment of meaning and understanding.

Ericsson and Simon (1981) have provided an excellent, concise summary of psychologists' views of cognition since the turn of the century. They begin by noting,

> The nature of data, and of admissible procedures for gathering and interpreting them, cannot be discussed very long in theory-free terms. The epistemological question (how the experimenter gains reliable information about the subject) is closely interwoven with the psychological question (how the subject gains and retains reliable information about the world). (p. 16)

> We have also long known, both from experiments and from everyday experience, how subjects' behaviors are affected by expectation, context, and measurement procedures. The notion that there can be "neutral" methods for gathering data has been refuted decisively. (p. 17)

The first method used by psychologists was what we would now call "naive introspection." Early psychologists such as William James considered it the major tool for psychological investigation. Binet defined psychology in terms of introspection. Titchner considered a phenomenological account, a description of consciousness, to be the starting point of a psychological description, but his aim was then to go beyond this and use the contents of consciousness to uncover its structure. In order to do this he tried to train his subjects to report only their immediate perceptual or imaginal experience, before that experience had been interpreted at all. He thought that in this way he could find the basic sensory units or attributes out of which all other experiences were then constructed.

The procedure was attacked from both sides. The Gestaltists claimed that the phenomenal experience was primary and could not be decomposed by analytic introspection. For them the immediate reports of naive subjects' experiences constituted the basic source of data. They emphasized the importance of conceptual understanding for learning and retention. The behaviorists, led by Watson, objected to the use of anything other than observable data. For Watson, the verbal reports, and not the events to which they supposedly referred, constituted the data. He cited the variability in the reports produced by different laboratories as evidence that such reports were inevitably contaminated by the subjects' prior experiences. He and many subsequent behaviorists thought that by sticking to observational data, they could avoid this problem and collect theory-free data.

Note that everyone relied on verbal reports as their primary data. The Gestaltists made extensive use of think-aloud procedures, but Watson did it first (Ericsson & Simon, 1981)—unless we count Freud, who had been having patients free-associate for years. The behaviorists and traditional personality assessors rejected verbal reports as evidence of direct experience, but starting in the 1970s cognitive assessors were once again asking clients to think aloud, and were accepting such reports as descriptive of the clients' phenomenal experience.

Cognitive therapies were first introduced by Ellis (e.g., 1962) and Kelly (1955), although not under that name, but did not gain even minimal acceptance until the 1970s. Mischel's (1973) article "Toward a Cognitive Social Learning Reconceptualization of Personality" is commonly cited as the turning point in the acceptability of cognitive formulations of personality.

An excellent current summary of *cognitive* assessment techniques can be found in a volume entitled, *Behavioral Assessment:* A

Practical Handbook (Parks & Hollon, 1988). These authors begin by noting the problems. Much cognitive processing takes place—and at any given time most thoughts, beliefs, and memories are—outside of conscious awareness. In contrast to earlier assumptions, it is now recognized that memory retrieval is largely a constructive, or reconstructive, process: "While people can often describe the products of their information processing, the actual processes followed may need to be inferred on the basis of carefully structured assessment paradigms" (p. 161).

Parks and Hollon (1988) describe instruments that are designed to measure thoughts, attitudes, beliefs, attributional styles, expectations, irrational thinking processes, and self-schemas. These assessment procedures are categorized as "recording," endorsement," "production," and "inferential" methods. Recording methods include listening to private speech (i.e., barely audible speech), especially of children, when they are unaware that their speech is being recorded; having people think aloud as they perform a task or solve a problem; asking people to free-associate; and asking people to respond to stimuli out loud. Endorsement methods involve presenting the subject with a set of items that have been either rationally constructed or empirically selected. Production methods include thought listing (listing one's thoughts in response to a stimulus), thought sampling (the same, except that thoughts are recorded in response to a beeper at random or fixed intervals), and event recording (recording one's thoughts during some designated event). Inferential methods are indirect in nature:

> The purpose of the assessment is not readily discernible from the procedure . . . Virtually all depend on inference; investigators do not ask subjects to report their operating self-schemata, personal constructs, etcetera. Rather, these cognitive structures are inferred from subjects' performance on a variety of cognitive and/or behavioral tasks, typically developed and validated in the basic cognitive psychology literature. (Parks & Hollon, 1988, p. 168)

"What's going on here?" readers may ask. How are these different from what personality assessors were doing 50 years ago, before they tried to get (pseudo)scientific? In one way or another people are being asked to state their thoughts, and those thoughts are either being accepted at face value or are being used to infer cognitive structures, beliefs, attitudes, and attributional styles. Sundberg (1981) has provided a fascinating history of cognitive assessment, showing that interest in thoughts and beliefs has been there all along. The inferences currently being made in cognitive assessment are more closely tied to the data than are the inferences of dynamic therapists. If someone suffers from a social phobia, the interest is in the thoughts that occur when the person is entering, or thinking about entering, the phobic situation. There may be some general theory concerning the kinds of thoughts that are most likely to lead to social anxiety (e.g., those having to do with evaluation by others), but the client's formulation of the belief is accepted, in contrast to inferring some repressed fear from early childhood as the real problem. Within the context of modern epistemologies it is no longer necessary to justify such procedures. It is sufficient to show that a persuasive argument can be made that the inferred relations make sense, or that the same kinds of thoughts occur in other people with similar problems.

Finally, before concluding this section I should note that there have been dramatic advances in cognitive assessment in its original sense. The computer has made possible the research program envisioned and attempted unsuccessfully by Galton, Cattell, and others just before and after the turn of the century. It now seems that it may be possible one day to measure intelligence in terms of neurological speed and efficiency (e.g., Glaser, 1981; Sternberg, 1981). With respect to both ability and content, cognitive assessment has completed the cycle to psychology's beginnings.

DEVELOPMENTS WITHIN THE TRADITIONAL MODEL

In this section, I describe three areas of research that represent developments within the traditional model, but which clearly reflect the influence of modern philosophies of science on that model. First, I describe some recent developments in the theory of personality inventory construction. Second, I make brief reference to a continuing effort to construct a comprehensive taxonomy of trait-descriptive adjectives. And, third, I briefly describe a very different approach to the construction of diagnostic indicators.

Personality Inventories

Multimethod Sequential Strategies

Jackson (1971) is usually credited with writing the manifesto for personality inventory construction in the 1970s and 1980s. He rejected the radical empiricist philosophy and called for an explicit concern with item content. Inventories were to be constructed within a theoretical framework; items written to measure constructs within that framework had to meet several criteria, including conceptual comprehensibility, to be included in scales. Jackson also argued that items should measure only one construct, should be free of social desirability, and should be keyed equally "true" and "false"—all criteria that I have already rejected.

Conceptually, the shift in thinking is indicated by the shift from criterion validation to construct validation (e.g., Embretson [Whitely], 1983; Messick, 1981). Whereas judgments or ratings would previously have been accepted as the criteria against which scales would be validated, it was now recognized that judgments and scales are both indicators of open rather than operationally defined constructs. Multitrait–multimethod matrices provide a systematic way of examining the relations among the indicators developed by different methods to measure a set of constructs. Divergent validity was seen to be as important as convergent validity, even if such data are not often provided in practice.

The shift in thinking is also reflected in the development and utilization of confirmatory factor analysis. Exploratory factor analysis had been used to discover the variables to be measured; now those variables can be derived on theoretical grounds, and confirmatory factor analysis can be used to determine the extent to which the instrument corresponds to the theory. Thus, empirical checks are still used, but in a very different way; for if the results do not fit the theory, it may well be the instrument rather than the theory that gets revised.

Finally, the concept of structural fidelity that Loevinger (1957) articulated long ago can comfortably be incorporated into the theory of inventory construction. The properties of the measurement model should mirror the properties of the construct of which it is an indicator. For example, a bipolar scale with "love" at one end and "hate" at the other is not appropriate for a theory including "ambivalence" (large amounts of love and hate existing simultaneously) as a construct. Such a theory requires separate, possibly orthogonal, scales for "love" and "hate."

Normative, Ipsative, and Idiothetic Strategies

Normative Strategies. Under the traditional model, most inventories have been constructed normatively. Scores on a scale indicate an individual's standing relative to a reference population on whatever the scale measures. All descriptors are relative to some (usually implicit) population or standard. With physical objects, the immediate measure, such as 6 feet, is relative to a standard; however, when we translate that measure into a descriptor via an ascription rule (Rorer & Widiger, 1982), the descriptor is relative to some population. Thus, 6 feet may indicate a tall woman relative to the general population of women, but a short man relative to other players in the National Basketball Association. When "very often" is used to describe the rate at which earthquakes occur, the rate is lower than when "sometimes" is used to describe the amount of shooting in a Western movie (Pepper, 1981). In psychology we do not have an objective standard such as a meter stick or a clock. Our measures are relative to start with, and are meaningful only when the reference population is specified.

The meaning of scores on normative personality scales is more basically relativistic than is always acknowledged. If we keep in mind that scales serve as indicators of constructs, and that the raw scores (which are meaningless) are translated into deviation scores relative to some population, the scores on two scales may indicate that an individual scores relatively higher on indicator A than on indicator B relative to this reference population; however, it does not necessarily follow that the individual has more of trait A than of trait B in any absolute sense, because that relation may change with a different reference population or with a different set of indicators for the same constructs.

Ipsative Strategies. Some personality inventories have been constructed ipsatively by the use of forced-choice items on which each of the possible responses is scored on a different scale. Each person has a fixed number of points, equal to the number of items, that can be divided up among the scales. Thus, every-

one's average raw score is the same, and the scaled results indicate the salience of a scale relative to the other scales for that person, *relative* to the reference population used in norming the scales. In this case it is the person's deviation relative to that person's other scores that is then normed relative to the reference population. As before, the indicated relation between two constructs might change with a different reference population or with a different set of indicators for the same constructs, and in this case it might also change if the two scales were compared with a different set of scales.

Idiothetic Strategies. In an article that should be a citation classic, but is not, Norman (1967) showed that the correlation between respondents' ratings of the social desirability of an item and their responses to the item could not be used to infer a relation between the ratings and the responses for an individual, because the estimator of the relation included population variance components. He even noted that in some published studies the *SDSV*s and the responses had been collected on different samples of respondents, thereby precluding any possibility that the correlation could indicate a relation at the individual level. He further showed that if one were willing to assume that the relation was constant for everyone, then the correlation between the interaction components (the doubly centered residuals with the main effects removed) provided an appropriate estimate of the within-individual relationship. A similar point was made by Gollob (1968a, 1968b) with respect to factor-analytic techniques.

More recently, Lamiell (1981, 1987) has advocated an alternative approach that he calls "idiothetic." Lamiell begins by noting the distinction, made by others, between personality psychology and individual-differences research. The latter he acknowledges to be appropriate for selection or classification purposes. However, as Norman and Gollob noted, individual-differences methods are inappropriate for drawing inferences concerning personality theory, which is concerned with relations that hold at the individual level for each individual.

Lamiell (1987) argues persuasively that the problem is rooted in the nomothetic–idiographic debate. Psychologists believed, and still believe, that the goal of personality research is to establish nomothetic laws—that is, laws that hold for each individual in the population—and that the way to do this is by means of individual-differences research. The latter is seen as appropriately objective and scientific, in contrast to idiographic research, which is believed to be subjective and therefore not scientifically respectable under the empiricist model. The problem is that individual-differences research cannot establish nomothetic laws, if they exist, of the kind that are desired.

According to Lamiell, the basic error is what William James called the "psychologist's fallacy": "This fallacy occurs whenever the empirical properties of *data* are uncritically assumed to reflect psychological properties of the *persons*[,] observations on whom generated the analysis of those data" (Lamiell, 1987, p. 91). He begins with an example with which few psychologists will have much trouble. If a poll shows that a population is split 51% to 49% on an issue, it does not follow that individuals do not have strong opinions on the issue—that is, that each individual's opinion is about 51–49% on the issue. For example, the fact that the population is fairly evenly divided on abortion rights does not mean that individuals do not hold strong views on the issue. Similarly, 90% or even 100% of the individuals may only slightly favor the same alternative. Thus, one cannot reason from the percentage split in the population to the strength of conviction of any individual.

The following is my formulation of the point that I understand Lamiell to be making. The basic fallacy is what logicians call the "fallacy of division" (Copi, 1982). The fallacies of composition and division are the reverse of each other, and they both involve a confusion between the distributive and collective uses of general terms. The fallacy of composition occurs when one argues from "attributes of the individual elements or members of a collection to attributes of the collection or totality of those elements" (Copi, 1982, p. 125). The fallacy of division occurs "when one argues from the attributes of a collection of elements to the attributes of the elements themselves" (p. 126). The *distributive* use of a term refers to each of the members or the collection individually; the *collective* use refers to the members all together, as a whole. For example, it may be that college students (distributively) may register for no more than six courses each term, but that college students (collectively) register for hundreds of courses each term.

Consider the following arguments, which are taken from Copi (1982). (1) Humans are mortal; Socrates is a human; therefore Socrates is mortal. (2) American Indians are disappearing; that man is an American Indian; therefore that man is disappearing. The first argument is valid, because "humans" is used distributively, that is, it applies to each member of the class. The second argument is invalid, because "American Indians" is used collectively, and therefore does not apply to each of the members of the class.

> The Fallacy of Division is committed when one argues that since a whole (or a class) has a specified attribute, therefore every part of that whole (or every member of that class) also has that attribute. Thus it would be an example of the Fallacy of Division to argue that since the United States is a wealthy nation, therefore United States citizen Jane Doe must be wealthy. It would also illustrate the Fallacy of Division to argue that because barbers have been cutting hair for hundreds of years, therefore the new barber at the corner barber shop must be very experienced. (Copi, 1982, p. 128)

Simply put, the correlations computed in individual-differences research are properties of the class, and applying them to individuals is an instance of the fallacy of division. There is no way to get from the relation between two traits or characteristics in the population to the relation between those traits within an individual. It follows, for example, that correlations between traits and behaviors in no way answer the question as to whether an individual with a particular trait is likely to exhibit a particular behavior. The latter is, in any event, a logical and not an empirical question (Rorer, 1989).

I now turn from Lamiell's critique of current research to his constructive suggestion for an alternative. Again, I am going to make the argument somewhat differently than he does, so there is the possibility that I am making a different argument. Personality researchers who have adopted the normative or nomothetic model have assumed that the meaning of terms such as trait descriptors is given by a person's standing relative to a reference population; for example, if we say that someone is average in friendliness it means that there are about as many people who are more friendly as there are people who are less friendly than that individual.

But there is another possibility. In making judgments about physical objects, people have explicit reference scales for characteristics such as height, weight, and temperature. There are also scales for hardness (of rocks) and intensity (of earthquakes). Perhaps what people do when they make judgments of other attributes is to use implicit scales of each of the attributes. Conceptually, these scales would range from none of the attribute to the maximum amount of it. As is the case with physical scales, the conceptual end points may not exist; for example, one may conceive of a scale from zero weight to infinite weight, even though there may be no objects with either zero or infinite weight, nor any way to measure them if there were. Lamiell calls this "idiothetic" measurement, and presents evidence that this, rather than normative measurement, may be what people have in mind when they make such judgments. Idiothetic measurement, if it can be implemented, will have the advantage of being relative to a standard rather than a population or other constructs.

Such a reconceptualization would also make sense out of several decades of rating scale research that psychometricians have found puzzling. The classic instance was the Navy's experience with scales labeled "poor," "fair," "average," "good," and "excellent." The scales were relabeled "average," "above average," "good," "very good," and "excellent" when it was found that the lowest rating ever given was "average." Similarly, at my university the teacher evaluation forms have a scale from 0 to 4, and the middle category, 2, is labeled "average." The average rating, however, is approximately 3.

It has been suggested that such results reflect leniency error, response bias, or some other form of rater irrationality. Another possibility—one closer to what students tell me about their thinking process in making the ratings—is that they have an implicit standard of, say, good teaching, and that they rate their professors relative to that scale. Relative to that scale, as opposed to relative to one another, their professors are viewed as above average (i.e., above the midpoint). The situation is obviously more complicated than that, because, as pointed out above in the discussion of ascription rules, people do make normative adjustments when making ratings. The debate about whether individuals make normative or idiothetic judgments when asked for personality ratings may be like the debate over whether persons add or average information in making

judgments: The answer is that sometimes they do one, sometimes they do the other, and often they do both.

However, it is important to distinguish the empirical question about what people do from the logical question about the appropriateness of inferences from data. Whatever studies may eventually show about how the average individual goes about making such judgments, three points seem clear: (1) Inferences concerning personality structure cannot be made from individual-differences data; (2) investigators should take seriously the possibility that respondents are making idiothetic, rather than normative, judgments, when interpreting their data; and (3) more research on the implementation of idiothetic measurement is sorely needed.

Taxonomic Approaches

I have indicated that factor-analytic research starts with the assumption that the ways in which individuals may differ psychologically are unknown, and attempts to establish empirically what those dimensions are. In order for a dimension to emerge, it must exist and it must be represented in the item pool. How can one know that all possible dimensions are represented in the item pool? One approach is to assume that any important differences among people will come to be represented in the natural language. If so, one could then start with the natural language and try to systematize the information that is included there. There is a long-standing and continuing effort to develop just such a taxonomy of trait-descriptive terms, and a description of those efforts would logically be a part of this chapter. John (Chapter 3, this volume) provides an excellent summary.

Taxometric Methods

I am going to return to the diagnostic model to describe briefly an area that would also require an entire chapter to describe adequately. Suppose that one is interested, not in dimensions along which individuals might differ, but in categories or classes of people that in some sense really exist; that is, the classes are nonarbitrary. Meehl (e.g., Meehl & Golden, 1982) has termed such classes "taxa," and the means of identifying them "taxometric methods." If one suspects that an entity (in our case specifically a diagnostic entity) exists, is there some way to determine which of the purported indicators of that entity are valid? The answer, surprising to those who have been schooled in traditional criterion-group research, is yes.

Let us start with the case where the criterion group is known to exist, but for some reason cannot be assembled—for example, people who will commit suicide or die of cancer. We want to know whether we can find an indicator that will identify these people in advance. Suppose we have two groups for which the suicide rate is known, even though we do not know which members of each group will commit suicide. If the suicide rates for the two groups are different, then it is possible to identify indicators (even perfect indicators) that will tell us which members of each group will commit suicide. Dawes and Meehl (1966) refer to this as "mixed-group validation."

Now consider an example such as schizophrenia, where it is not known whether any such disorder exists in the taxonomic sense, or if there may be more than one disorder. Given a sizable sample, one can determine the optimum cutting point on a potential indicator by noting the covariance between two other potential indicators for various regions of the first indicator. The optimum cut on the first indicator is in the region of that indicator for which the covariance between the other two indicators is a maximum. The optimum cutting point on the other two indicators can then be found in similar fashion. Given a set of indicators for which optimum cuts have been determined, Bayesian probabilities of taxon membership can be determined for each individual in the sample on the basis of that person's score on each indicator. Details are given in Meehl and Golden (1982).

Both of these procedures are examples of what have come to be known as "bootstrap procedures," by which it is meant that marginally valid indicators can be used to identify more valid (even perfectly valid) indicators. For example, clinical judgments might be used to form two groups that differ in suicide rate. Suppose that 50% of those whom a clinician identifies as suicide risks subsequently attempt suicide, whereas only 10% of those who are identified as nonrisks do so. These groups can now be used to identify more valid indicators of suicide risk (Dawes, 1967). This may seem paradoxical at first, but should not if one considers the history of medicine. Today's sophisticated diagnostic techniques have all been developed on the basis of the crude indicators that existed a couple of hundred years ago.

NONTRADITIONAL MODELS

Postempiricist Philosophies of Science

The demise of empiricism ended the search for a line of demarcation between science and metaphysics (nonscience); removed the requirement for justification of knowledge claims; and opened the way for the return of the study of cognition, consciousness, and meaning to psychology. Scientists are still expected to be able to defend their positions with good reasons and strong arguments, but modern philosophers are more humble, more aware and acknowledging of our limitations. It is accepted that we can never know anything for sure. All knowledge is filtered through our perceptual apparatus, and there is no way to know how that apparatus distorts the information we receive. Given that absolute knowledge is impossible, all we can ask is whether we have good reasons for believing what we do.

The early positivists limited meaningful statements to reports of sense data. The later empiricists accepted the meaningfulness of statements concerning physical objects that were independently verifiable, but the ontological question concerning the reality of physical objects was still considered to be metaphysical. Without the empiricists' requirement for justifiable knowledge, it is possible to embrace realism on the basis of reasonableness. Somewhat ironically, as the changing philosophy of science has made it easier for psychologists to hold a realist position, fewer of them have done so; instead, they have embraced a constructivist position that emphasizes the arbitrary nature of psychological entities (e.g., Gergen, 1985; Hampson, 1988). There is a middle ground. Although I acknowledge the logical problems, I believe that one can reasonably argue for an ontological realism while holding a pragmatic, constructivist epistemology. Given this view, constructs are admittedly constructed, but reality, which we cannot know directly, places limits on the extent to which different constructions will work. Those that work, we keep. With respect to psychological constructs, in particular, there are probably many that will work to varying degrees.

In a recent article Kukla (1989) outlined some of the nonempirical issues to be dealt with in psychology. Many of these would provide no problem for an empiricist. Theory construction, for example, was always considered to be in the context of discovery rather than the context of justification. The philosophers certainly considered their own analytical work to be appropriate. The development of theoretical systems by means of logical deduction was a highly esteemed activity, as was the analysis of empirical theories for conceptual coherence. In many cases theoretical analysis can result in new knowledge. Conceptual innovations allow us to understand existing knowledge in a different way. Copernicus's heliocentric theory is an example of a conceptual innovation that included no new empirical information, but facilitated the development of other theories explaining the movements of the heavenly bodies.

Some philosophers believe that there may be other fundamental issues involving a priori knowledge and basic presuppositions that are not empirically reconcilable. Behaviorists (discussed above) and phenomenologists (to be discussed below) do not disagree about the use of introspective reports; they disagree about what those reports mean. Resolving the differences between them is not an empirical matter. It involves examining the presuppositions of the two positions. Kukla (1989) suggests that psychology may not need more data as much as it needs conceptual innovations concerning the data we already have. This chapter suggests that a good case can be made that this is true of personality assessment.

With this background I now describe two overlapping approaches that fall clearly in the postempiricist era—and, as we shall see, in the pre-empiricist era as well.

Narrative Analysis: Hermeneutics

Cognitive assessment, in my view, differs fundamentally from behavioral assessment in its incorporation of meaning as an important aspect of human functioning. Cognitive models of personality and personality dysfunction focus on beliefs and on the logic of an individual's thinking. The behaviorists used simplified experimental situations (e.g., Skinner boxes) and nonhumans, and so eliminated much of the role that meaning plays in determining human

behavior. In the simplified experimental situation, the stimulus was clear; in real life, it is not. An individual may perceive me as a man for census purposes, a man in the social classification sense, a man in the sexual sense, a professor, his or her professor, a father, his or her father, a lover, a psychologist, or a psychotherapist. What I am as a stimulus to another person is not an objective fact that can be known by observation.

Likewise, the fact that the response in the box is defined in terms of its effect rather than the behavioral topography does not make a lot of difference, because in that setting one can ignore the question of whether the animal intended to press the lever, whereas in ordinary human affairs the question of intent is crucial. If someone else's hand comes in contact with me, it makes a great deal of difference whether this person is giving me a friendly greeting, congratulating me, teasing me, hitting a mosquito, punching me, or just swinging his or her arm; and, again, this is not an objective fact that can be determined by observation. It is something that must be determined contextually, and even then the person may mislead me as to his or her intent.

The behaviorist focuses on the relation between a stimulus and a response and attempts to explain the latter in terms of the manipulation of the former. The cognitive theorist notes that the same objective stimulus may produce different response topographies in different people (e.g., a grade of B may cause some to leap in glee and others to weep), and attempts to explain those differing responses in terms of the intervening cognitions of the people. The behaviorist would argue that, in principle, the explanation can be found in the person's prior history without resorting to cognitions. For the cognitivist, the meaning of the stimulus to the perceiver and the intent (meaning) of the response to the responder are crucial. They use the methods I have described to try to determine those meanings; as we have seen, those methods are the methods of traditional personality assessment.

A growing number of theorists believe that these methods are insufficient and inappropriate for the study or understanding of meaning. The traditional methods of natural science were designed for the study of material things, and meaning is not a material thing; it is an emergent property. Just as water has emergent properties that cannot be predicted from the properties of oxygen and hydrogen, and life has emergent properties that cannot be predicted from a knowledge of the properties of the elements of which living organisms are constituted, meaning is a further property that has emerged as a consequence of, or along with, consciousness.

Meaning is not a thing or substance; it is an activity. Because each of us has direct access only to our own experience of meaning, the realm of meaning is difficult to investigate, cannot be investigated by natural science methods, and has therefore not been the object of much study in psychology. Meaning is expressed linguistically, and linguistic data can be understood only by means of hermeneutic reasoning or understanding.

"Hermeneutics" originally referred to the study and understanding of the Bible, but was then broadened to include the study and understanding of all texts. Now, some psychologists believe that human experience is constructed in the same way as, and therefore must be understood in the same way as, a text. Understanding cannot come in the form of general laws from which deductions are made, but only in terms of the life history—the narrative that represents the person's experience. Individual or personality psychology differs from differential or individual-differences psychology. It is represented by studies of lifespan development, such as case histories and biography, for which the appropriate methodology is hermeneutics.

Polkinghorne (1988), on whose exposition the following is based, uses the term "narrative" to refer broadly to the attempts that we make to understand our actions and those of others by creating narratives (accounts, stories, scenarios, etc.). If asked why I am typing at my computer, I would explain by describing my commitment to write this chapter and my desire to finish it quickly, given that I am already far past the deadline to which I originally agreed. Most people would accept this as an appropriate explanation of my actions, on the grounds that I have made my actions understandable by giving them meaning—something that can not be experimentally or objectively determined. There is a reason why I am typing rather than playing golf or watching television.

Experience is neither material nor static; it is constantly changing as we create narratives to make sense out of what we are perceiving and thinking. Our understanding of various actions and events may change as the result of additional experiences. To return to an earlier example, I may have been angry when I thought the other person was trying to hurt me when he or she hit me, but when I learn that the other person was killing a mosquito that was biting me my anger turns to appreciation, because I have rewritten the narrative. The construction of experience does not proceed logically; it proceeds sequentially, associationally, pictorially, and emotionally. Logical thought is (unfortunately) but a small part of the narrative construction and experience.

One of the empiricists' goals was to construct a language that would permit objective communication. It is now acknowledged that this cannot be done. We learn language contextually. As we learn new words, our understanding of old words changes. The meaning of a sentence changes, depending on the context in which it is used. And so with stories and the larger context in which they are embedded. Hermeneutics makes this contextualism explicit. Texts must be analyzed in relation to our knowledge of the author and the times when the author wrote, and in relation to our own intuitive understanding of human experience.

Psychoanalysts' use of the case study was an early attempt at a hermeneutic approach to individual psychology. Although Freud's theory evolved as he attempted to explain new cases, psychoanalysis was an attempt to find a representation that would work for all people for all time. A hermeneutic analysis would emphasize the turn-of-the-century European Jewish culture from which Freud and his colleagues came, the sexual repressiveness of Western society in general, the mechanistic theories of physics and the knowledge of anatomy and physiology at the time, Freud's own life history, and so forth. Such an analysis would facilitate our understanding of psychoanalysis, and would lead to the conclusion that psychoanalysis is but one way in which to understand human behavior. To the extent that psychoanalysis imposes a construction, a personologist working from a hermeneutic approach would probably not find it very useful, because that investigator would want to understand the constructions of the person being studied.

Murray's (1938) *Explorations in Personality* reports a later effort to understand individuals on the basis of detailed information about them. After that, academic interest in biographical study virtually disappeared from psychology until the 1970s, when works such as *The Seasons of a Man's Life* (Levinson, 1978) began to appear, and there was increasing interest in the study of lifespan development. More recently there has been an increasing number of articles on qualitative research in general, and on narrative analysis and hermeneutics in particular. For representative examples, one might try recent papers by Mishler (e.g., 1979) or Rosenwald (e.g., 1988). Gergen, Hepburn, and Fisher (1986) applied hermeneutics to the process of personality assessment as it is currently practiced, with devastating results. They found that items on the Rotter Internal-External Scale "could plausibly be used as indicators of virtually any common trait term within the English language" (p. 1261). So much for the idea that an item can be thought to represent just one trait.

Phenomenological Perspectives

Some will think that this should have been the major section of this chapter; others will wonder why it is here at all. Such is the divided state of personality assessment.

We have seen that at the turn of the century there was great interest in finding true or justifiable knowledge. The interest is understandable. The church had for centuries claimed to be the arbiter of true knowledge, and the rationalist philosophers had attempted to provide a philosophical basis for that claim in pure reason. The empiricists launched a challenge that carried the day into the middle of this century. The arbiter of true knowledge was science, not the church. Justifiable knowledge came from experience, not from mere, meaningless, metaphysical speculation.

The empiricists had originally sought true knowledge by limiting their data to sense data, but there were problems with that position. From a philosophical point of view, it was open to the charge of being solipsistic, and so it was revised to require intersubjective testability. It was not until the second half of this century that it was realized that the structuralists' program of introspection at the turn of the century should have shown on psychological grounds that sense data were unacceptable as an epistemological criterion, because it was impossible to free them of the effects of prior learning (Rorer, in press).

At the turn of the century there was still

another claim that sense data were the correct road to true knowledge. Husserl's efforts provided the foundation of phenomenology (Jennings, 1986), which was swept aside by the rising tide of empiricism. Husserl did not reject the use of empirical methods to gain true knowledge of the material world, but he argued that this solved only half of the problem. The other half was to understand the process of knowing. Because consciousness is a process rather than a material thing, it cannot be studied by methods appropriate to studying things. Husserl sought to provide a means whereby one could obtain absolute knowledge of conscious phenomena, something he felt was essential if psychology was to understand how naive experiential reports are contaminated by preconceptions. Before a psychologist can study the relation between a stimulus and a perception, it is important to know what the perception is, and phenomenological methods were designed to provide that information. The basic idea was that one could learn to put aside one's preconceptions and experience the essence of an experience—that part of the experience that would be equally true at any other time and place (i.e., in any other context). Essences cannot be established by reason or induction, but must be immediately experienced. The means of achieving this experience was the reduction, the purpose of which is to move from psychological consciousness to pure consciousness—from a concern with empirical facts to a concern with essences. These are inseparable. The act of knowing and that which is known are not independent. For a more complete description, see Jennings (1986), on whose excellent exposition my account is based.

As with any school of knowledge, phenomenology has evolved and has developed different subschools (Henley, 1988; Valle, King, & Halling, 1989). The belief that one can acquire true or absolute knowledge has been abandoned by all except religious fanatics. Phenomenologists would now acknowledge that essences cannot be grasped independently of preconceptions. The aspects on which Snyder (1988) believes they would agree are the following:

1. the importance of experience in its own right, rather than as a reflection of an independently existing reality;
2. the emphasis on intentionality (i.e, the directedness of consciousness towards something, or alternatively, that consciousness is consciousness of something) as the fundamental structure around which the exploration of experience can occur, as opposed to the consideration of psyche in terms of mental contents that like physical entities do not have a referential aspect;
3. the importance of describing this experience in as clear and direct a manner as possible;
4. through emphasizing essence, the importance of understanding the fundamental nature of whatever is being investigated;
5. the route toward this understanding and subsequent description is directly given through intuition and is not fundamantally dependent on any other epistemological avenue (e.g., the application of inductive reason to experimental data). (p. 404)

Clearly, someone with this orientation is going to have a very different view of personality assessment than any of those whom we have considered so far. Many phenomenologists do not do assessments, because they believe that assessment necessarily treats the person as object, neglecting the person as subject. Fischer (1989) is an exception. She terms her approach a "human science" approach, in contrast to the prevailing "natural science" approach. The emphasis is on considering the person as a whole, and on the differentiations that can be made within that whole, in contrast to the normative framework of traditional assessment. The individual's experiences, not the underlying traits indexed by standard tests, are considered primary. Linear causal explanations are rejected in favor of descriptions of the whole.

> Within human science, both research and applied work begin with different data than that which mainstream psychology has accumulated. Like behaviorists, we do begin with publicly observable events. . . . However, in contrast, we regard these events as implying the person's lived history, present world, and goals. . . . We do use test scores, traditional research, and other theories' constructs to help us expand and refine our understandings of those observed events. But we regard these resources as tools, ones derived from earlier life events. (Fischer, 1989, pp. 164–165)

> The assessor recognizes that the client's experience of his or her situation must be taken into account. The assessor's own experience also is inevitably involved. . . . Disciplined, examinable descriptions, then, are achieved not by trying to exclude the assessor's subjectivity (individuality, perspective, involvement) but by explicitly taking it into account. (Fischer, 1989, p. 165)

SUMMARY AND CONCLUSIONS

Because all forms of personality assessment presuppose, and are embedded in, a philosophy of science, changes in the philosophical *Zeitgeist* have strongly influenced personality assessment. One cannot understand the historical development of personality assessment unless one puts it in the context of changes in the philosophy of science. In the late 19th century empiricism and phenomenalism, among others, competed with religion to be recognized as the path to true knowledge. The ascendance of empiricism in philosophy set the course of psychology for 50 years. The extremes of empiricism were replaced during the 1960s by an equally extreme rejection of empiricism. Now we see more moderate positions that acknowledge the impossibility of identifying true knowledge, or of distinguishing science from nonscience. There is increasing interest in the late-19th-century pragmatism that was among the casualties swept aside by empiricism.

Paralleling these developments in philosophy, first empirical and then behavioral assessment rose in favor, while cognitive, phenomenalistic, and rational approaches diminished. But then positions that had been temporarily eclipsed by the dominance of logical empiricism in psychology were again seen as viable alternatives, and positions that flourished during the heyday of empiricism have quietly faded away. Rational considerations have returned to personality inventory construction as brute-force empirical procedures have given way to a demand for conceptual coherence between measures and theories. As cognitive behaviorists have challenged methodological behaviorists to acknowledge that it is the construal or evaluation of an object or event, rather than the object or event itself, in terms of which behavior is best explained, cognitive assessments have supplanted or replaced behavioral assessments.

A growing number of researchers believe that the meaning an individual attaches to objects or events can be understood and explained only in the context of that person's life, and that the methodology appropriate for the study of lives is narrative analysis or hermeneutics. There has been a corresponding growth in the number of textbooks and classes devoted to or including qualitative as well as quantitative research methods. Introspectionism, the lessons from which should have precluded the rise of empiricism, has not returned as the means of studying cognitive structure, but there is growing interest in phenomenological reports as a means of understanding personality. Unlike the traditional case studies that attempted to fit every individual into a single psychodynamic framework, modern cognitive and phenomenological assessments focus on the individual's framework for understanding his or her life.

Personality assessment is not a unified enterprise; it is a divided camp. Many traditional personality assessment researchers find themselves uncomfortable with the current changes. If the rigid rules of objectivity and operationism do not hold, does that mean that "anything goes"? No. The current positions are a far cry from the world-views position, popular in the 1960s, which held that any representation of reality is as good as any other. Alternate views are possible (e.g., wave and particle theories of light, cognitive and behavioral explanations of behavior), but not any view (e.g., demonological theories of behavior). However, there is no firm line of demarcation between acceptable and unacceptable views. There may well be many different ways of representing the world, with no way to say that one is correct and the rest false; however, it is reasonable to believe that there is a world out there, and that some representations will therefore work better than others. To keep the ones that work and discard the ones that do not is to adopt a pragmatic position with which most scientists should find themselves comfortable. The judgment as to which ones work best, or have the best hope of working best, is not a judgment that can be made objectively, so reasonable people may reasonably disagree.

It is important to take seriously the distinction between research on individual differences and research on personality theory. The individual-differences structure of the population is not the same as the personality structure of the individual. Evidence concerning the dimensions on which individuals differ does not provide an explanation for the behavior of those individuals. Most of what is included in traditional personality assessment would better be described as individual-differences assessment.

The data and methods of analysis that are

appropriate for identifying individual differences are not the same as those that are appropriate for developing personality theory, and the differences may run deeper than has been realized. So far, the efforts that have been made have been to reconcile the two approaches—to adapt individual-differences methods to idiographic ones by means of idiothetic procedures. That approach is creative and deserves support. But it may be that the ultimate resolution of much of the tension in the field will come from realizing that traditional personality assessment procedures are appropriate for the study of individual differences, whereas behavioral, cognitive, hermeneutic, and phenomenological methods are appropriate for developing personality theory.

ACKNOWLEDGMENTS

Preparation of this chapter was supported in part by National Institute of Mental Health Grant No. MH-39077.

Bobbie G. Hopes provided advice and encouragement throughout the preparation of this chapter. William H. Henricks at Miami University, and Lewis R. Goldberg, Sarah E. Hampson, Oliver P. John, and Tina K. Rosolack at Oregon Research Institute responded with incredible speed to my urgent plea for assistance under great time pressure. The chapter benefited greatly from their comments and suggestions. My thanks to all.

REFERENCES

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

Bentler, P. M., Jackson, D. N., & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin, 76,* 186–204.

Berg, I. A. (1967). The deviation hypothesis: A broad statement of its assumptions and postulates. In I. A. Berg (Ed.), *Response set in personality assessment* (pp. 146–190). Chicago: Aldine.

Bleuler, E. (1950). *Dementia praecox or the group of schizophrenias* (J. Zinkin, Trans.). New York: International Universities Press. (Original work published 1911)

Block, J. (1965). *The challenge of response sets: Unconfounding meaning, acquiescence and social desirability in the MMPI.* New York: Appleton-Century-Crofts.

Block, J. (1971). On further conjectures regarding acquiescence. *Psychological Bulletin, 76,* 205–210.

Burisch, M. (1986). Methods of personality inventory development—comparative analysis. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaires: Current issues in theory and measurement* (pp. 109–120). New York: Springer.

Butcher, J. N. (Ed.). (1987). *Computerized psychological assessment: A practitioner's guide.* New York: Basic Books.

Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology, 56,* 754–761.

Copi, I. M. (1982). *Introduction to logic* (6th ed.). New York: Macmillan.

Dawes, R. M. (1962). A note on base rates and psychometric efficiency. *Journal of Consulting Psychology, 26,* 422–424.

Dawes, R. M. (1967). How clinical judgment may be used to validate diagnostic signs. *Journal of Clinical Psychology, 23,* 403–410.

Dawes, R. M., & Meehl, P. E. (1966). Mixed group validation: A method for determining the validity of diagnostic signs without using criterion groups. *Psychological Bulletin, 66,* 63–67.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.* New York: Dryden Press.

Ellis, A. (1962). *Reason and emotion in psychotherapy.* New York: Lyle Stuart.

Embretson (Whitely), S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93,* 179–197.

Ericsson, K. A., & Simon, H. A. (1981). Sources of evidence on cognition: A historical overview. In T. V. Merluzzi, C. U. Glass, & M. Genest (Eds.), *Cognitive assessment* (pp. 16–51). New York: Guilford Press.

Fischer, C. T. (1989). Personality and assessment. In R. S. Valle & S. Halling (Eds.), *Existential–phenomenological perspectives in psychology* (pp. 157–178). New York: Plenum.

Gergen, K. J. (1985). The social constructionist movement in modern psychology. *American Psychologist, 40,* 266–275.

Gergen, K. J., Hepburn, A., & Fisher, D. C. (1986). Hermeneutics of personality description. *Journal of Personality and Social Psychology, 50,* 1261–1270.

Gilberstadt, H., & Duker, J. (1965). *A handbook for clinical and actuarial MMPI interpretation.* Philadelphia: W. B. Saunders.

Glaser, R. (1981). The future of testing: A research agenda for cognitive psychology and psychometrics. *American Psychologist, 36,* 923–936.

Goldberg, L. R. (1971). A historical survey of personality scales and inventories. In P. McReynolds (Ed.), *Advances in psychological assessment* (Vol. 2, pp. 293–336). Palo Alto, CA: Science and Behavior Books.

Goldberg, L. R. (1972). Parameters of personality inventory construction and utilization: A comparison of prediction strategies and tactics. *Multivariate Behavioral Research Monographs, 72*(2).

Goldberg, L. R., & Slovic, P. S. (1967). Importance of test item content: An analysis of a corollary of the deviation hypothesis. *Journal of Counseling Psychology, 14,* 462–472.

Goldstein, G., & Hersen, M. (Eds.). (1984). *Handbook of psychological assessment.* New York: Pergamon Press.

Gollob, H. F. (1968a). Confounding of sources of variation in factor-analytic techniques. *Psychological Bulletin, 70,* 330–344.

Gollob, H. F. (1968b). A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika, 33,* 73–115.

Hampson, S. E. (1988). *The construction of personality: An introduction* (2nd ed.). London: Routledge.

Harré, R., & Lamb, R. (1983). *The encyclopedic dictionary of psychology.* Cambridge, MA: MIT Press.

Hathaway, S. R., & Meehl, P. E. (1951). *An atlas for the clinical use of the MMPI.* Minneapolis: University of Minnesota Press.

Hayes, S. C. (1986). Behavior philosophy in the late 1980's. *Theoretical and Philosophical Psychology, 6,* 39–43.

Henley, T. B. (1988). Beyond Husserl. *American Psychologist, 43,* 402–403.

Jackson, D. N. (1971). The dynamics of structured personality tests: 1971. *Psychological Review, 78,* 229–248.

Jennings, J. L. (1986). Husserl revisited: The forgotten distinction between psychology and phenomenology. *American Psychologist, 41,* 1231–1240.

Kazdin, A. E. (1979). Fictions, factions, and functions of behavior therapy. *Behavior Therapy, 10,* 629–654.

Kelly, G. A. (1955). *The psychology of personal constructs* (2 vols.). New York: Norton.

Kraepelin, E. (1981). *Clinical psychiatry* (7th ed.) (A. R. Diefendorf, Trans.). Delmar, NY: Scholars' Facsimiles & Reprints. (Original work published 1907).

Kukla, A. (1989). Nonempirical issues in psychology. *American Psychologist, 44,* 785–794.

Lamiell, J. T. (1981). Toward an idiothetic psychology of personality. *American Psychologist, 36,* 276–289.

Lamiell, J. T. (1987). *The psychology of personality: An epistemological inquiry.* New York: Columbia University Press.

Levinson, D. J. (1978). *The seasons of a man's life.* New York: Ballantine Books.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3* (Monograph No. 9), 635–694.

Marks, P. A., & Seeman, W. (1963). *The actuarial description of personality: An atlas for use with the MMPI.* Baltimore: Williams & Wilkins.

McReynolds, P. (1986). History of assessment in clinical and educational settings. In R. O. Nelson & S. C. Hayes (Eds.). *Conceptual foundations of behavioral assessment* (pp. 42–80). New York: Guilford Press.

McReynolds, P., & Ludwig, K. (1987). On the history of rating scales. *Personality and Individual Differences, 8,* 281–283.

Meehl, P. E. (1982). Subjectivity in psychoanalytic inference: The nagging persistence of Wilhelm Fliess's Achensee question. In J. Earman & C. W. Savage (Eds.), *Minnesota studies in the philosophy of science: Vol. 10. Confirmation of theories* (pp. 349–411). Minneapolis: University of Minnesota Press.

Meehl, P. E. (1986). Trait language and behaviorese. In T. Thompson & M. D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 315–334). Hillsdale, NJ: Erlbaum.

Meehl, P. E., & Golden, R. (1982). Taxometric methods. In P. Kendall & J. Butcher (Eds.), *Handbook of research methods in clinical psychology* (pp. 127–181). New York: Wiley.

Messick, S. (1981). Constructs and their vicissitudes in educational and psychological measurement. *Psychological Review, 89,* 575–588.

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review, 80,* 252–283.

Mishler, E. G. (1979). Meaning in context: Is there any other kind? *Harvard Educational Review, 49,* 1–19.

Murray, H. A. (1938). *Explorations in personality.* New York: Oxford University Press.

Norman, W. T. (1963). Relative importance of test item content. *Journal of Consulting Psychology, 27,* 166–174.

Norman, W. T. (1967). On estimating psychological relationships: Social desirability and self-report. *Psychological Bulletin, 67,* 273–293.

Parks, C. W., Jr., & Hollon, S. D. (1988). Cognitive assessment. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (3rd ed., pp. 161–212). New York: Pergamon Press.

Paulhus, D. L. (1986). Self-deception and impression management in test responses. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaires: Current issues in theory and measurement* (pp. 143–165). New York: Springer.

Paulhus, D. L. (in press). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. Wrightsman (Eds.), *Measures of personality and social-psychological attitudes.* New York: Academic Press.

Pepper, S. (1981). Problems in the quantification of frequency expressions. In D. W. Fiske (Ed.), *Problems with language imprecision* (pp. 25–42). San Francisco: Jossey-Bass.

Polkinghorne, D. E. (1988). *Narrative knowing and the human sciences.* Albany: State University of New York Press.

Rorer, L. G. (1965). The great response style myth. *Psychological Bulletin, 63,* 129–156.

Rorer, L. G. (1989). *Misconceptions concerning states and traits.* Unpublished manuscript.

Rorer, L. G. (in press). Some myths of science in psychology. In D. Cicchetti & W. Grove (Eds.). *Thinking clearly about psychology: Essays in honor of Paul E. Meehl.* Minneapolis: University of Minnesota.

Rorer, L. G., Hoffman, P. J., Laforge, G. E., & Hsieh, K. C. (1966). Optimum scores to discriminate groups of unequal size and variance. *Journal of Applied Psychology, 50,* 153–164.

Rorer, L. G., & Widiger, T. A. (1982, May). *Ascription rules for trait descriptive adjectives.* Paper presented at the meeting of the Midwestern Psychological Association, Minneapolis.

Rorer, L. G., & Widiger, T. A. (1983). Personality structure and assessment. *Annual Review of Psychology, 34,* 431–463.

Rosenwald, G. C. (1988). A theory of multiple-case research. *Journal of Personality, 56,* 239–264.

Snyder, D. M. (1988). Comment on Jennings. *American Psychologist, 43,* 403–404.

Sternberg, R. J. (1981). Testing and cognitive psychology. *American Psychologist, 36,* 1181–1189.

Sundberg, N. S. (1981). Historical and traditional approaches to cognitive assessment. In T. V. Merluzzi, C. U. Glass, & M. Genest (Eds.), *Cognitive assessment* (pp. 52–76). New York: Guilford Press.

Suppe, F. (1977). *The structure of scientific theories* (2nd ed.). Urbana: University of Illinois Press.

Valle, R. S., King, M., & Halling, S. (1989). An introduction to existential–phenomenological thought in psychology. In R. S. Valle & S. Halling (Eds.), *Existential–phenomenological perspectives in psychology* (pp. 3–16). New York: Plenum.

Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment.* Reading, MA: Addison-Wesley.

PART V
OVERVIEW

2

Why Use Personality Tests? A Brief History and Some Comments

John E. Exner, Jr.
Philip Erdberg

Although writers and poets have struggled for centuries with the question of what makes each person unique, the notion of personality, as a subject for psychological study, is less than 100 years old. Allport (1937) noted that a movement called the "psychology of personality" gained momentum only after 1920, yielding numerous but conflicting theories and plentiful but inconsistent research. The works of Freud had, of course, received considerable attention beginning in the late 1800s. But they failed to attract those attempting to address the challenge of defining personality, and it is Allport who can be credited with bringing the issue into sharper focus. He reviewed some 50 definitions or descriptions of personality and neatly summarized their underlying commonality by defining personality as the dynamic organization, within the individual, of those psychophysical forces that determine his or her unique adjustment to his or her environment.

In formulating this definition, Allport also noted that the uniqueness of each individual created havoc for a psychological science whose emphasis was on articulating universal laws of human behavior. He put forth a very compelling challenge: namely, that psychology would achieve its true purpose only when it could deal with the issue of individuality. Allport also offered a strong caution about research orientations that presented only fragments of information about the total person and then attempted to extend that information at the expense of individual differences.

Shortly after Allport's classic *Personality* was released, Henry Murray published another classic, *Explorations in Personality* (1938). He did not attempt to define personality as Allport had done, instead approaching the issue of individuality from a different direction. He emphasized it through case studies that described the unique integration of various characteristics within individuals. These studies neatly illustrated how information derived from many sources, including psychological tests, could be used to develop a fine-grained picture of a person. Both Murray and Allport distinguished between idiographic and nomothetic approaches to the study of people and, in their own way, each argued for an integration of the two. They advocated an approach that would not simply judge a person against others but would contrast a person's unique features with those found in others.

The arguments put forth by Allport and Murray stimulated thinking about the objectives of

the psychological study of people and, in effect, created a challenge for those interested in testing or assessment. The idea of assessing personality and psychopathology began to be of greater interest to psychologists, although they often used the two terms interchangeably. And, early on, most attempts to use tests to understand people actually did not include much personality testing, focusing instead on issues of intelligence, aptitude, achievement, and vocational interests. A few psychologists, such as Robert Woodworth, had attempted to devise methods to identify psychopathology around the time of World War I. Some of those efforts persisted into the 1920s and 1930s with the development of inventories designed to measure traitlike features such as introversion, extraversion, and neuroticism, but none was used extensively in clinical settings. Issues of individuality typically were addressed with data drawn from histories and interviews.

It was not until the late 1930s that Allport and Murray's approach to understanding people as individuals through the use of multiple data sources moved into the clinical mainstream. It mushroomed at an almost incredible pace during and immediately after World War II with the huge expansion of mental health services provided by the military and the Veterans Administration. The test battery became a standard component of clinical practice and was probably best elaborated by the Rapaport group at the Menninger Foundation during the 1940s, leading to the procedure that came to be called psychodiagnosis.

The objective of psychodiagnosis involved much more than settling on a diagnostic label. Psychodiagnosis was a multitest procedure designed to study the person as a unique entity. Implicit in the process was the premise that an understanding of the person's assets, deficits, and conflicts could contribute in some significant way to his or her therapeutic well-being. By describing those organized and recognizable elements and suggesting how they led to behavior, including the presenting symptoms, the psychodiagnostic process could help to inform a treatment plan.

Throughout the 1940s and 1950s, clinical psychologists became well recognized and highly regarded for their psychodiagnostic expertise and the input they provided for treatment planning conferences. But unfortunately, those interested in the study of personality began to move toward one of two increasingly polarized positions. One, more empirically oriented, argued for the study of individual traits and the relation each had to behavior. The second group agreed with many of the basic ideas of the trait theorists but argued strongly that traits could not simply be conceptualized as a group of independent characteristics, each of which was either present or absent in an individual. They preferred to think of personality as a unitary entity and argued that a trait is a descriptive, nonexplanatory concept that, following from Allport and Murray, must be weighed in terms of its strength or importance for the individual's overall economy. Conversely, those arguing for the study of traits as independent characteristics emphasized the empirical aspects of their methodology but maintained that their approach had the potential for ultimately representing personality as an integrated entity.

The disagreement about how best to approach issues of personality and individual differences became more complex during the early 1950s, when a falling out occurred between those who argued strongly in favor of a so-called objective testing approach and a second group that became closely aligned with what was known as the projective psychology movement. The former argued for a more nomothetic approach to the study of personality, while the latter defined itself as more oriented toward understanding the unique individual. Unfortunately, no one profited from the schism, and it was not uncommon in those days for members of one group to avoid those in the other. And although psychodiagnosis—or personality assessment as it began to be called—continued to be an important part of the clinical psychologist's role, it did so with far less uniformity than had previously been the case.

During the late 1950s, another force emerged, with significant implications for clinical psychology practice. It was the advent of radical behaviorism, which brought with it the

notion of the black box—the message that there is no such thing as personality—and the assertion that, even if there is, it certainly could not be measured through psychological testing. This movement created a new group of clinicians who not only avoided personality assessment but campaigned very actively against it, favoring instead the tactics of observation and counting of critical incidents as ways of determining intervention targets. Thus, by the early 1960s, the once reasonably homogeneous specialty of personality assessment had fragmented considerably and had begun to come under serious attack.

These developments probably had their greatest impact during the mid 1960s and through most of the 1970s, as clinical psychology training programs changed considerably in regard to personality assessment. Students often objected to the laborious training time required to learn about test batteries and their applications. Their protests were frequently reinforced by both clinical and nonclinical faculty members, who voiced opinions that the specialty of psychodiagnosis or personality assessment had only limited value. Many programs reduced training in personality theory and personality assessment, moving instead to a technique-oriented emphasis on treatment. It is sad but true to note that, for at least a decade, these programs produced graduates who knew little about assessment and less about personality, and yet tended to extol their newly acquired therapeutic expertise.

The assumption in these new programs was that psychological tests created only restricted and nongeneralizable samples of performance, not allowing for the kind of reliable and valid personality description that would usefully cut across a variety of situations. Instead, their focus became the identification of manifest symptom constellations—borderline personality, panic disorder, anorexia, and posttraumatic stress disorder serve as examples—with the idea that, once identified, these constellations called for specific intervention approaches. This unreasonable logic neglected the individual as a unique entity and, even more important, was based on the naive assumption that the presentation of a psychological symptom could dictate a specific therapeutic technique This medical-model approach in which symptom identification can lead automatically to treatment, while sometimes appropriate for conditions such as appendicitis, clearly is not well suited for working with individuals in ways that exemplify responsible mental health practice.

An illustration may be useful in demonstrating how personality assessment can uniquely differentiate patients who present similar symptom pictures. The emphasis personality assessment places on understanding people as individuals ensures that each will receive the most appropriate treatment. The illustration is taken from two real-life cases, both women between the ages of 27 and 32, each of whom complained of intensely disruptive bouts of anxiety with frequent panic attacks. Similar in their presentations, if they had consulted psychiatrists specializing in psychopharmacology, treatment would probably have consisted of some kind of antianxiety medication; if either had consulted a psychologist specializing in anxiety or panic reactions, tension reduction and stress management would have been the likely intervention. Fortunately, however, each was referred for personality assessment before treatment was prescribed.

The assessments done by two psychologists moved well beyond the presenting problem to demonstrate how two very different women could have arrived at similar symptom pictures. In each case, personality assessment helped to describe current and long-term resources, self-concept and interpersonal issues, and how each managed affect and cognition.

The psychologist responsible for the first case reported that Ms. A works as a secretary in an accounting firm and was recently separated after four years of marriage. She was quite defensive in spite of the fact that she has considerable resources, with stress tolerance and capacity for control usually as robust as those of most adults. However, these assets were currently less effective because of some situationally related stress, probably associated with her recent separation. The effects of the stress were relatively modest in absolute terms, but they had

created a state of psychological overload. This overload appeared to have produced a potential for impulsiveness that was more likely to manifest in her emotional displays than in her thinking. Much of her stress appeared related to an experience of emotional loss and probably translated as feelings of loneliness or neglect. Ms. A emerged as an intuitive person who typically involves feelings in her psychological operations. However, at present those operations had become more emotionally complex and intense than was typically the case, leading to feelings of confusion and being out of control.

The psychologist also noted that Ms. A is an intensely self-centered person who frequently overestimates her self-worth. She is apt to focus much more on herself than on others, and her interpersonal relationships may often be tenuous and nonmutual. She is likely to view rejections as insults to her unrealistically skewed self-concept, one that is based more on imagination than a balanced understanding of herself in relation to others. Nonetheless, it is important for her to defend her self-esteem, and she typically externalizes responsibility, especially for negative or unpleasant events.

Ms. A also has a strongly dependent orientation, frequently leaving her in a hostile-dependent bind in relation to others. She would like to be in unconditionally supportive and nurturing relationships, but unfortunately this increases her vulnerability to the manipulations of others. She appears to have some awareness of this bind, and she is considerably less secure about her interpersonal relations than are most people. She attempts to conceal or contend with these feelings of uncertainty by frequently using an intellectual, somewhat authoritarian approach. Although she is open to social interaction, she is cautious and sometimes even reluctant to initiate interpersonal exchanges, especially those that may require tact or maturity. She is especially defensive about relationships that may create unwanted demands or pose hazards to her control of the situation.

As noted above, Ms. A is an intuitive person whose feelings play an important part in the decisions she makes. She is not reluctant to display those feelings, and she may often convey the impression of being emotional or even impulsive. Her thinking is usually clear, but her current stress state tends to interfere with her abilities to concentrate or attend to specific events. This appears mostly to be a situational issue, since the assessment data suggested that she has no major problems in reality testing and seems as prone as most adults to make conventional responses when the circumstances of the situation clearly define expected or acceptable answers.

Personality assessment helped to describe Ms. A as a person with substantial resources who was presently feeling overwhelmed. Intensely emotional, highly self-focused, and vulnerable to interpersonal rejection, she was likely to experience anxiety and panic in response to a situation in which her usually effective coping operations could no longer provide adequate control. It is a picture very different from the one described by the psychologist who evaluated Ms. B for similar symptoms.

Ms. B, who works as a costume designer for a theatrical company, emerged on personality assessment as a very conservative and cautious person who seems quite insecure about herself and her ability to deal effectively with the world. She is especially reluctant to deal with ambiguity or complexity, and she has developed an orientation toward coping and decision making in which she tries to keep things at a simple, easily managed level.

Although she is a very ideational person, she devotes much of her thinking to fantasy as a way of avoiding the stresses of reality. And indeed, she *is* vulnerable to disorganization even in the context of routine stress because she has fewer coping resources readily available than do most adults. Consequently, she is conservative in processing new information and seems especially fearful that her feelings could disorganize her ability to make decisions. She tries to avoid emotionally charged situations whenever possible, particularly those in which there is a more than superficial interpersonal component.

Personality assessment depicts Ms. B as a somewhat psychologically impoverished, fragile individual who tries to simplify a world she finds overly complex. Although she is interested in

people, her understanding of them is based much more on fantasy than on actual experience. Thus, although she would enjoy more closeness, she seems bewildered about how best to achieve it and concerned about what sacrifices she might be called upon to make in return. It is likely that she lives in an ongoing state of anxiety and apprehensiveness, punctuated by episodes of panic when situations occur that threaten her limited emotional resources.

Psychodiagnostic assessment suggests that Ms. A and Ms. B. took very different roads to their similar symptom presentations of anxiety and panic. By describing each as an individual—not as a set of symptoms—it becomes possible to formulate intervention strategies that have the greatest likelihood of being effective.

The psychologist reporting on Ms. A's findings presented a convincing argument for a supportive intervention to assist in working through the current stress situation created by her recent separation. The findings also highlighted how important it would be to alert the treating therapist to her strongly passive-dependent orientation, which has significant implications for the therapeutic relationship. The psychologist also identified Ms. A's tendency to externalize blame and her sometimes volatile expression of feelings as possible treatment targets. But the report noted that those issues likely would not be open for intervention unless a more insight-oriented therapy were to evolve from the initial supportive approach. It also described the assets that Ms. A would bring to treatment: considerable psychological resources, a relatively consistent coping style, and energetic and accurate information processing.

The recommendations that emerged from Ms. B's personality assessment were very different. The psychologist argued that her anxiety and panic were functions of her precarious existence—one in which she is never sure whether her attempts to simplify the world will protect her from stress. When she is viewed as a unique person, intervention objectives are easy to identify, but a strategy for achieving them is more difficult to define than it was for Ms. A.

First, the findings suggested that Ms. B must be approached very cautiously. She has significant deficits in her understanding of herself and others. Her resources are limited and her abusive use of fantasy serves only to prevent her from acquiring more adequate coping skills. She is the type of person who, if confronted with the need for some long-term form of treatment, would likely bolt because the prospect would be so threatening in its complexity. It would be better to broach the treatment issue in a more specific way that also leaves the door open for future developments as appropriate. This could involve suggesting that treatment begin with a focus on broadening social skills and contending more directly with feelings as ways of easing anxiety and panic symptoms. The report also alerted the treating therapist to Ms. B's avoidant and oversimplifying style and her frequent retreats into fantasy, longstanding coping strategies that a therapist might otherwise view as resistance to treatment.

Even though their presenting symptoms were similar, Ms. A and Ms. B emerged on personality assessment as very different people. Only an extreme optimist could believe that the same treatment would benefit both. Ms. A appeared to need supportive intervention that might evolve into a longer-term, insight-oriented engagement of some of the characterological issues that emerged. Ms. B requires a more developmentally oriented approach whose focus would be the significant deficits that assessment revealed.

These illustrations emphasize the importance of personality assessment in understanding the individual more thoroughly and planning treatment more precisely. Some have contended that personality assessment is time consuming and unnecessary, supporting their position with the questionable argument that the same information will ultimately emerge as treatment progresses. Actually, the procedures involved usually take no more than a few hours and significantly increase the likelihood that the person will move into the most beneficial treatment, a particularly important concern in an age of short-term interventions. No one would go into surgery or begin some other form of medical intervention without first being assured that the relevant tests had been completed and

that the attending physician had a good understanding of all the issues and treatment alternatives. People seeking mental health treatment deserve the same standard of care.

It is, however, less than certain that this level of practice will be routinely available in an era of managed care. A recent summary article by Piotrowski (1999) evaluated the impact of managed health care policies on psychological assessment. He noted that although psychologists continue to value traditional assessment approaches, they are less likely to use them because of the constraints imposed by managed care. As an example, Piotrowski, Belter, and Keller (1998) found that 72% of respondents in a *National Register* sample were doing less assessment overall than 5 years ago and that they had decreased their pool of assessment instruments, moving toward brief, symptom-focused self-report measures. Groth-Marnat (1999) summarized these trends, noting that, "Many managed health care organizations (MCOs) view assessment, especially through test batteries, as being an unnecessary frill and question the extent to which it either accelerates or optimizes client outcome" (p. 782).

It is true that many issues concerning the assessment of personality remain open to inquiry, but it is well-designed research that has the best potential for addressing the concerns posed by managed care. Most notably, research is needed to provide nomothetic support for the kind of person-oriented assessment illustrated in the case studies above. Such research would, for example, study the ability of personality assessment to predict response to specific intervention models. That kind of research would fit with the increasing demands managed care programs place on therapists to support their decisions with data-based rationales, and it could actually mandate the use of assessment in treatment planning. Fisher, Beutler, and Williams (1999) noted that the increased requirement for psychotherapists to justify their interventions ". . . has forced clinicians to reconsider the traditional practice of viewing all patients through the same theoretical lens. Instead, clinicians must plan treatment differentially—matching each patient to the treatment that has the greatest probability of success" (p. 826).

REFERENCES

Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt, Rinehart & Winston.

Fisher, D., Beutler, L. E., & Williams, O. B. (1999). Making assessment relevant to treatment planning: The STS clinician rating form. *Journal of Clinical Psychology*, 55(7), 825–842.

Groth-Marnat, G. (1999). Current status and future directions of psychological assessment. *Journal of Clinical Psychology*, 55(7), 781–785.

Murray, H. A. (1938). *Explorations in personality*. New York: Oxford University Press.

Piotrowski, C., Belter, R. W., & Keller, J. W. (1998). The impact of managed care on the practice of psychological testing: Preliminary findings. *Journal of Personality Assessment*, 70, 441–447.

Piotrowski, C. (1999). Assessment practices in the era of managed care: Current status and future directions. *Journal of Clinical Psychology*, 55(7), 787–796.

4.12
Principles of Personality Assessment

JERRY S. WIGGINS and KRISTA K. TROBST
University of British Columbia, Vancouver, BC, Canada

4.12.1 INTRODUCTION 350

4.12.2 PSYCHODYNAMIC TRADITION 350
4.12.2.1 Historical Background 350
4.12.2.2 Conceptual Framework 351
4.12.2.3 Assessment Instruments 351
4.12.2.4 Interpretive Principles 351
4.12.2.4.1 Projective hypothesis 352
4.12.2.4.2 Levels of functioning 352
4.12.2.4.3 Psychological adjustment 352
4.12.2.5 Applications and Current Status 352

4.12.3 PERSONOLOGICAL TRADITION 353
4.12.3.1 Historical Background 353
4.12.3.2 Conceptual Framework 353
4.12.3.3 Assessment Instruments 354
4.12.3.4 Interpretive Principles 354
4.12.3.5 Applications and Current Status 355

4.12.4 MULTIVARIATE TRADITION 355
4.12.4.1 Historical Background 355
4.12.4.2 Conceptual Framework 356
4.12.4.3 Assessment Instruments 356
4.12.4.4 Interpretive Principles 356
4.12.4.5 Applications and Current Status 357

4.12.5 EMPIRICAL TRADITION 357
4.12.5.1 Historical Background 357
4.12.5.2 Conceptual Framework 358
4.12.5.3 Assessment Instruments 359
4.12.5.4 Interpretive Principles 359
4.12.5.4.1 Empirical strategy 359
4.12.5.4.2 Constructive strategy 359
4.12.5.4.3 Interpersonal strategy 359
4.12.5.5 Applications and Current Status 360

4.12.6 INTERPERSONAL PARADIGM 360
4.12.6.1 Historical Background 360
4.12.6.2 Conceptual Framework 361
4.12.6.3 Assessment Instruments 361
4.12.6.3.1 Interpersonal Adjective Scales 361
4.12.6.3.2 Inventory of Interpersonal Problems 361
4.12.6.3.3 Impact Message Inventory 362
4.12.6.4 Interpretive Principles 362
4.12.6.5 Applications and Current Status 363

4.12.7 CONCLUSION 364
4.12.8 REFERENCE 365

4.12.1 INTRODUCTION

Consider a hypothetical clinician who is both well trained and well informed about available personality tests. In her daily practice, she employs a range of assessment instruments (singly or in combination) to address the range of diagnostic issues posed by her clients. When "thought disorder" is suspected, she might occasionally administer a Rorschach and from most clients she will obtain a case history. Her clients themselves may be asked to fill out the NEO-Personality Inventory to clarify dimensions of normal personality functioning or the Minnesota Multiphasic Personality Inventory-Revised (MMPI-2) to assess gross abnormality. She might also routinely administer the Interpersonal Adjective Scales to assess the character and quality of clients' interpersonal relationships.

Like many of her colleagues, our clinician may describe her theoretical orientation as "eclectic" and indeed that is what she appears to be. But she is also, perhaps unknowingly, operating within larger historical contexts of assumptions and beliefs that may include traditions of thought as diverse as psychoanalysis, psychobiography, psychometrics, empirical realism, and symbolic interactionism (respectively for the tests just mentioned). We will refer to such contexts as different "traditions" of personality assessment, by which we mean the background of a set of unquestioned beliefs or orienting attitudes within and against which personality tests are administered and interpreted.

The traditions that will be considered in this chapter are referred to as psychodynamic, personological, multivariate, empirical, and interpersonal. Such traditions are related, but *not* equivalent, to certain theories of personality (Hall & Lindzey, 1978) and psychopathology (Millon, 1969). Similarly, these traditions are related, but *not* equivalent, to certain kinds of personality tests. Thus, the MMPI has been associated with a distinctive philosophy of science (empirical realism) rather than with a particular theory of psychopathology. The Rorschach test and the Wechsler scales have figured prominently within the ego psychological perspective of the psychodynamic tradition, although both tests were constructed for quite different purposes.

We believe that these traditions reflect different and potentially useful ways of viewing the entire psychodiagnostic enterprise. We also believe that these traditions have contributed to more or less invisible barriers to communication being erected among clinicians who work, knowingly or not, within them. This state of affairs is unfortunate because "membership" in a tradition, as it were, is mainly determined by where one went to graduate school, with whom one studied, and the particular clinical setting in which one finds oneself operating. Further, we are convinced that, at the present time, there are some remarkably similar ideas being advocated about the nature of personality and its assessment among those working within several different traditions, and that recognition of these commonalties might advance the field as a whole (see Wiggins, in press).

We would venture to suggest that most readers will have familiarity with a few of these traditions, but not with some of the others. For example, those intimately familiar with the psychodynamic tradition are less likely to be familiar with the multivariate and empirical traditions, and vice versa. Those who are fond of obtaining life history data from interviews may not be fully aware of the rich history of the personological tradition nor of the exciting recent developments that have occurred within that tradition. And many may not be aware that the interpersonal tradition has fostered a strong "underground movement" within clinical psychology for many decades.

4.12.2 PSYCHODYNAMIC TRADITION

4.12.2.1 Historical Background

David Rapaport (1911–1960) was a major systematist of psychoanalytic theory, the originator of a now standard psychodiagnostic test battery, and a highly influential mentor of the principal architects of the psychodynamic tradition in personality assessment. After receiving his PhD from the Royal Hungarian University, he emigrated to the USA in 1938 and shortly thereafter joined the staff of the Menninger Clinic where he eventually became chief psychologist and head of the research department. During and shortly after World War II, the results of an extensive program of collaborative research were summarized in a two-volume *Manual of diagnostic psychological testing* (Rapaport, 1944–46) that eventually "revolutionized clinical psychology and influenced clinical psychologists the world over" (Gill & Klein, 1967, p. 18).

Rapaport moved to the Austin Riggs Center in 1948 and continued his extensive collaborations with colleagues at institutions such as the Menninger Foundation, Yale University Department of Psychiatry, and the Research Center for Mental Health at New York University. Speaking collectively for workers at these and other institutions, Roy Schafer (1967) observed, "All of us are working within the psychoanalytic psychodiagnostic tradition crystallized by David Rapaport" (p. 2).

4.12.2.2 Conceptual Framework

It is difficult to make a coherent set of distinctions among the many "psychodynamic" theories that have been proposed as alternatives to Freud's classical psychoanalytic theory (Westen, 1990). It would seem to be generally true however that since, and even before, Freud's death, there has been an increased emphasis on the autonomous and adaptive functions of the ego and on the internal representation of self and others within the ego domain. The aspect of classical psychoanalysis that has been de-emphasized, or even abandoned by some, is the proposition that personality and its development are largely determined by unconscious libidinal and aggressive drives.

Rapaport (1960) must be counted among the most rigorous systematists of classical psychoanalytic theory and he was also a major contributor to the alternative perspective known as "ego psychology." In both his theoretical and assessment work, his principal interest was in thought processes and their course of development from drive-oriented to reality-oriented processes. In this context, Hartmann's (1939) concept of the independence (from drive) of the ego and its role in fostering active processes of adaptation were appealing to him, as was Erikson's (1950) emphasis on the cultural and historical contexts in which the ego develops. Rapaport (1960) developed and expanded the metaconcept of structure which he felt was the "prime requisite for progress toward dimensional quantification" (p. 98) and which he applied broadly to many aspects of psychological functioning and cognitive organization (see Gill & Klein, 1967).

The generally increased emphasis upon ego functions in postclassical psychoanalysis was also reflected in the currently important psychodynamic alternative of object relations theory. Proponents of this view challenged the classical idea that cathexes of external "objects" (including persons) served mainly as vehicles through which instinctual energies were discharged. Instead, it was postulated that early interactions with significant others ("objects") led to internalized representations of both others (Jacobson, 1954) and self (Kohut, 1971) that serve as "internal working models" (Bowlby, 1973) for later interpersonal relationships. From this perspective, the person is, from birth (Main, Kaplan, & Cassidy, 1985), an object seeker (Fairbain, 1952) who establishes a "gratifying involvement" with other persons (Behrends & Blatt, 1985).

The shift in emphasis within psychoanalytic theory from drives to internalized representations has been accompanied by a corresponding shift in rationales for personality assessment, as exemplified by the ego-psychological approach of Allison, Blatt, and Zimet (1968) and the object-relations approach of Blatt and Lerner (1983).

4.12.2.3 Assessment Instruments

Prior to Rapaport's influential contributions in the 1940s, assessment psychologists were primarily technicians who administered IQ tests. Since that time, they have become clinicians who administer batteries of both structured and projective tests of personality and cognition. The multitest battery was advocated in view of the apparent complexity of personality and cognition and their interrelated functions, as well as for the purpose of gathering normative data that would shed light on those complexities (Rapaport, Gill, & Schafer, 1946). At Menninger, the composition of the battery reflected judgments regarding the potential of each instrument to yield measures that might be interpreted within the ego psychological framework of Rapaport and his associates.

The projective component of the original Menninger battery included the Rorschach Test (Rorschach, 1921), the Thematic Apperception Test (TAT; Morgan & Murray, 1935), and a locally constructed Word Association Test. The nonprojective component included the Bellevue Scale (Wechsler, 1941), the Babcock Test of mental efficiency (Babcock, 1933), the Sorting Test of concept formation (Goldstein & Scheerer, 1941), and the Hanfmann–Kasanin test of concept formation (Hanfmann & Kasanin, 1937). As this test battery evolved over a 20-year period, the Rorschach, TAT, and WAIS (Wechsler, 1958) became the more or less standard core of the psychodynamic test battery (e.g., Allison et al., 1968).

4.12.2.4 Interpretive Principles

The vast literature on interpretive principles associated with the psychodynamic paradigm

defies easy summarization. Rapaport's theoretical work (e.g., Gill, 1967) and the writings of those whom he influenced (e.g., Holt, 1967) are, understandably, highly technical in nature. Of greater help to the clinician are the case summaries (Allison et al., 1968; Prelinger & Zimet, 1964; Rapaport et al., 1946; Schafer, 1948) and the individual treatments of the Rorschach (Schafer, 1954), TAT (Holt, 1951), and WAIS (Blatt & Allison, 1981). The following statements regarding the projective hypothesis, levels of functioning, and psychological adjustment are meant only to convey some of the flavor of three, among many, important interpretive principles.

4.12.2.4.1 Projective hypothesis

The projective hypothesis states that "All behavior manifestations of the human being, including the least and the most significant, are revealing and expressive of his personality, by which we mean that individual principle of which he is the carrier" (Rapaport, 1942, p. 92). Thus, people's possessions—clothes, automobiles, furniture—are expressive of their personalities and reflect single acts of choice; in their totality, they reflect the organization of such choices. Responses to the ambiguous stimuli of projective tests may also be thought of in terms of "choice," although such choices are much less conscious or volitional in nature. Thus, responses to a Rorschach inkblot may be thought of as reflecting "choices" between forms, colors, shadings, and so forth, to which a subject imparts meaning through organization. Responses to a TAT card also involve both choice (e.g., which figure to identify with) and organization (e.g., sequence of events). Responses to intelligence and concept formation tests involving choice and organizational processes may be used as "nonprojective tests of personality," given an adequate theory of "functions underlying the reactions and achievements on these tests" (Rapaport, 1946, p. 228).

4.12.2.4.2 Levels of functioning

The psychoanalytic model of primary (pleasure principle) and secondary (reality principle) modes of thought is meant to account for both a developmental sequence and characteristics of the mature adult (Rapaport, 1951). Consequently, there is a continuum of adult psychological functioning that may be assessed with an appropriate battery of tests.

> This continuum ranges from functioning in situations which put a premium on highly logical, reality-oriented secondary modes of thought (WAIS) to those which allow for more personal, less conventionally constrained thinking (TAT) and finally those which allow for considerably novel, personalized, and regressive modes of thinking (Rorschach). (Allison et al., 1968, p. vii)

Assessment of level of functioning has been greatly facilitated by Holt's innovative procedures for assessing primary and secondary process in the Rorschach (Holt & Havel, 1960).

4.12.2.4.3 Psychological adjustment

The Rorschach, TAT, and WAIS may be employed to assess both adaptive capacities and impairments in psychological functioning and to identify the functions impaired in different psychiatric diagnostic groups. Adjustment and maladjustment may be assessed with reference to the following postulated sequence:

> Certain patterns of defense mechanisms are adopted and these determine specific strengths and weaknesses in psychological functioning which then become characteristic of the adjustment of the personality; with the onset of maladjustment, an exaggeration or breakdown in these strengths and weaknesses characteristic for that maladjustment occurs which can be measured; this leads to a diagnostic differentiation. (Rapaport, Menninger, & Schafer 1947, p. 249)

4.12.2.5 Applications and Current Status

Since the late 1960s, theory and method within the psychodynamic paradigm of personality assessment have evolved into an object relations perspective that is highly compatible with contemporary formulations of social cognition, information processing, attachment research, and ego development (see Westen, 1990). Sidney Blatt and his associates at Yale have been primarily responsible for this paradigm shift. From the early 1950s until his untimely death in 1960, Rapaport contributed to the development of ego psychology and its application to personality assessment; since the early 1970s, Blatt has contributed to the development of object relations theory and its application to current personality assessment methods.

"The study of the representation of the human form on the Rorschach is an ideal data base for assessing an individual's representational world—his conception of people, including himself, and their actual and potential interactions. The representation of people, that is, object representations, have both structure and content" (Blatt & Lerner, 1983, p. 8). The structural aspects of object representations are emphasized in the Rorschach scoring system developed by Blatt, Brenneis, Schimek, and Glick (1976) that provides a developmental analysis of object representations in terms of such categories as differentiation, articulation, and integration of object and action. The content and affective themes of object representations are emphasized in the Rorschach scoring system developed by Mayman (1967) that emphasizes phenomenological dimensions such as affect states, ego states, experience of self, and sense of identity.

An object relations scoring system for the TAT has also been developed by Westen (1991) in which stories are rated for complexity of representations of people, affective tone of relationship paradigms, capacity for emotional investment, and understanding of social causality. A similar scoring system has been developed for the Picture Arrangement subtest of the WAIS (Westen, 1991, p. 72). In conclusion, the three major instruments of the original Menninger battery continue to show promise under a revised conceptual orientation that is most compatible with current thinking in personality, social, clinical, and developmental psychology.

4.12.3 PERSONOLOGICAL TRADITION

4.12.3.1 Historical Background

The "storied" nature of human conduct (Sarbin, 1986) has been recognized for many centuries within the disciplines of history, literature, and most of the social sciences. Within the much more recent personological tradition in personality assessment, the person is taken to be the basic unit of observation and the person's life story is considered to be the preferred way in which the person is to be understood (McAdams, 1994). The origins of the personological tradition, and in fact the origins of academic personality psychology itself, can be traced most directly to Harvard University in the 1930s. At that institution, Gordon Allport (1937) wrote the first textbook on personality, which defined the person as the basic unit of observation. Subsequently, he emphasized the use of personal documents in psychobiography (Allport, 1942, 1965) and spent a lifetime pondering the question of "How shall a psychological life history be written?" (Allport, 1967). At the Harvard Psychological Clinic, Murray (1938) introduced a multiform organismic method for studying "lives in progress" (White, 1952), in which an interdisciplinary team of investigators applied a variety of assessment procedures to the study of individuals over the course of their lives. During World War II, a variant of this procedure was employed in the selection of intelligence agents and saboteurs (Office of Strategic Services, 1948), and subsequently this approach spawned a variety of successful peacetime assessment programs (e.g., MacKinnon, 1975; Stern, Stein, & Bloom, 1956).

Murray's enduring contributions to the personological paradigm go well beyond those just mentioned. During his 60 years at Harvard, Murray was a source of personal inspiration for an extraordinarily talented and diverse group of students, colleagues, and visitors who, over the years, have applied and expanded his "personology" up until the present time. Early associates at the clinic included Samuel Beck, Erik Erikson, Jerome Frank, Daniel Levenson, Donald MacKinnon, Silvan Tomkins, and Robert White. Dan McAdams and William McKinley Runyan are more recent products of this Harvard tradition.

4.12.3.2 Conceptual Framework

Murray's (1938) personology provided a set of orienting attitudes that guided the multiform (many assessors, many tests), organismic (holistic) approach to the study of lives over time (e.g., White, 1966, 1975). Murray's (1959) conceptual framework was meant to serve as "the scaffold of a comprehensive system," rather than as a completed formal theory. Nevertheless, Murray (1938) introduced a number of concepts that have subsequently proven useful in guiding the study of lives. By taking the life cycle of the individual as the largest unit of study, Murray was committed to studying personality "the long way" (White, 1981), and he introduced units of time varying in duration and complexity from the proceeding (single episode) to the unity-thema (central life motif). A thema may be understood in terms of the interaction of needs (for which Murray developed his famous taxonomy) and press (environmental facilitation or obstruction of a need). The TAT (Murray, 1943) was developed as a projective measure of need–press interaction and has been utilized in large-scale studies of needs for achievement (McClelland, 1961), power (Winter, 1973), and intimacy (McAdams, 1989).

Erik Erikson's (1950) psychodynamic theory of personality development provided a more explicit account of the stages of psychosocial development over the entire life span (from the original psychosocial crisis of "trust versus mistrust" to the final crisis of "ego-integrity versus despair"). His theory informed his own classic psychobiographies, such as those of

Martin Luther (Erikson, 1958) and Gandhi (Erikson, 1969), as well as contributing to more recent psychobiographical efforts (e.g., Stewart, Franz, & Layton, 1988). More recently, in a book entitled *The stories we live by: Personal myths and the making of the self*, McAdams (1993) presented a neo-Eriksonian theory of identity development that is enriched by more contemporary concepts such as image, prototype, and script, and that is guided by the metatheoretical concepts of agency and communion (Bakan, 1966). This work has been hailed as "the most original and important new book on personality theory since George Kelly's *Psychology of personal constructs*" (Hogan, 1994, p. 356). The heuristic potential of McAdams' conceptual framework is suggested by the quality and quantity of critical response it has elicited from personality psychologists representing diverse theoretical perspectives (see McAdams, 1996 and peer commentary therein). Although McAdams' framework is more formal, explicit, and testable than earlier efforts, its roots in the formulations of Allport and Murray are easily traceable (McAdams, 1994).

4.12.3.3 Assessment Instruments

In what is now widely recognized as the definitive general treatise on life histories and psychobiography, Runyan (1982) observed:

> It should be clear that there is no single life history method, any more than there is a single personality research method, and that life histories may be studied through phenomenological self-reports, archival research, prospective longitudinal research, and experimental research. (p. 6)

Personality assessment psychologists from Murray to McAdams have recognized the need for methodological pluralism (Craik, 1986) in the study of lives, and for that reason the "boundaries" of the personological paradigm have historically been somewhat more permeable than those of the other paradigms considered in this chapter. Nevertheless, it would be inaccurate to characterize this paradigm as "eclectic," because the diversity of methods employed reflects the different levels of analysis encompassed by the term "life history," rather than the selective use of theories and methods to "explain" a given life. Following Kluckhohn and Murray (1953), Runyan (1982) distinguished three levels of generality in the social sciences: the universal (general laws of human behavior); the group (differences in sex, social class, culture); and the individual (distinctive, distinguishing characteristics of a particular person) and surveyed attempts to establish acceptable criteria for interpretations at each level. Although personologists in general, and clinicians in particular, are primarily concerned with the individual, an awareness of the larger enterprise that Runyan portrays serves to clarify the nature and goals of individual personality assessment. And for those few who might seriously be considering becoming psychobiographers, Alan Elms' (1994) book is essential reading.

The 25 different assessment procedures employed for the study of individuals at the Harvard Clinic (Murray, 1938) were indeed "multiform" in terms of both methods and the assessors who employed them; for example, autobiography (Murray), hypnosis (White), level of aspiration (Frank), dramatic productions (Erikson), and Rorschach (Beck). Perhaps of greatest interest to the contemporary clinician is the case study method in which some form of interview is employed with all the attendant advantages and limitations of retrospective, introspective, and qualitative methods (see Runyan, 1982, chap. 8). The procedures employed by McAdams (1993, chap. 10) provide an excellent example of a theory-driven approach to the interview.

4.12.3.4 Interpretive Principles

The variety of personality research methods for studying an individual's life history is reflected in a similar variety of interpretive principles associated with different methods. In this context, the interpretive principles elucidated by Alexander (1990) would appear to be "mainstream" in the sense that they evolved directly from the framework of Murray and his co-workers (especially Tomkins 1947, 1979) at the Harvard Psychological Clinic.

In analyzing data from directed interviews or autobiographical essays, Alexander (1990) looks for recurring dynamic sequences (scripts, themas, guiding messages) that may be revealed by "letting the data speak" or by "asking the data a question." In the former method, the typically large data set may be reduced by identifying "significant" material according to the following nine criteria of salience: (i) primacy (that which occurs first), (ii) frequency (that which occurs often), (iii) uniqueness (that which is unusual or odd), (iv) negation (that which is denied or disavowed), (v) emphasis (that which is either overemphasized or underemphasized), (vi) omission (that which is missing by normative standards or by implication), (vii) error or distortion (factual errors and Freudian slips), (viii) isolation (that which does not "fit" and non sequiturs), and (ix) incompletion (that which is not finished or lacks closure).

The method of "asking the data a question" is similar to that employed in the analysis of TAT protocols for predetermined categories such as "power motivation" (Winter, 1973). The database of interview material is reduced by selecting every sequence or incident related to the questions posed by the investigator. When applied to interview material, this method retains the "projective" advantage of addressing issues that the storyteller did not consciously intend to describe or deal with, while circumventing problems such as the identification of the "hero" of the narrative (Alexander, 1990).

4.12.3.5 Applications and Current Status

In recent years, there appears to have been a "back to basics" movement in personality assessment research (Wiggins & Pincus, 1992) that has rekindled interest in the fundamental assumptions of several paradigms, including the personological: "Once, again, it is okay to study the 'whole person.' Better, contemporary personologists insist, as did pioneers like Gordon Allport, that such an endeavor is the personologist's *raison d'être*" (McAdams, 1988, p. 1).

In terms of clinical applications, it should be borne in mind that the personological tradition has tended to focus on relatively normal, nonclinical samples and that within this tradition, life histories have been viewed as relatively veridical accounts of what has happened in a person's life (Runyan, 1982) or as imaginative reconstructions of the past (McAdams, 1993). It is also important to recognize that personality assessment within the personological paradigm occurs on a different level than traditional psychodiagnostic assessment (Alexander, 1990). Whereas personological assessment attempts to understand current personality functioning in terms of an individual's own life history, traditional psychodiagnostic assessment is concerned with determining the individual's membership in a group or class of individuals that differs from other groups in terms of psychopathology.

> A major difference between the two approaches is that the results of the former [personological] can easily be directed toward answering the questions intended by the latter [psychodiagnostic]. The reverse is, unfortunately, a low probability event in anything other than a global sense. To designate someone as obsessive or hysteric, or depressive, or schizophrenic will place the focus on particular salient aspects of functioning but say little about the dynamics in that individual leading to that particular form of functioning. (Alexander, 1990, p. 7)

4.12.4 MULTIVARIATE TRADITION

4.12.4.1 Historical Background

Sir Francis Galton's (1888) method for analyzing the "co-relations" between twin pairs was refined by Pearson (1896) and utilized effectively by Spearman (1904) in identifying a "general factor" of intelligence (*g*) that appeared to "underlie" various tests of mental abilities. Factors are statistical abstractions that summarize the relations among test scores in terms of underlying or "latent" variables which may be thought of as operating at several levels. For example, the intelligence quotient (IQ) could be considered to be a general higher-order factor that summarizes relations between verbal and performance major group factors, which in turn, summarize relations among multiple minor group factors (Vernon, 1950).

Raymond B. Cattell, the founding father of the multivariate tradition of personality assessment, was a student of Spearman and was among the first to apply factor-analytic methodology to the study of temperament (e.g., Cattell, 1933). Using cluster-analytic procedures, Cattell (1943) reduced Allport and Odbert's (1936) exhaustive compilation of trait-descriptive terms from Webster's unabridged dictionary to 35 "surface traits" (see John, 1990). He administered the 35 surface-trait variables (e.g., talkative vs. silent) to respondents in a series of peer-rating studies and eventually concluded that no less than 16 "source traits" could account for the interrelations among the 35 surface clusters (Cattell, 1949).

The Guilfords (1936, 1939) were also pioneers in the application of factor analysis to personality data, and they assembled evidence for the four interpretable factors of "shyness," "rhathymia," "depression," and "liking for thinking" which would now be called " extraversion," "conscientiousness," "neuroticism," and "intellect," and which (together with "agreeableness") constitute the currently important five-factor model (FFM) of personality (Digman, 1996). A decade later, Eysenck (1947) factored the intercorrelations among presenting symptoms of psychiatric patients and found two factors (extraversion and neuroticism) that had both theoretical (Jung, 1971) and empirical (MacKinnon, 1944) precedents (see Eysenck, 1990, pp. 95–98). By the 1950s, the systems of Eysenck (1953), Cattell (1957), and Guilford (1959) had become extensive alternative theories of personality structure.

The history of the multivariate tradition is a curious one that has recently been reconstructed

by Digman (1996). Briefly, Digman argues that as early as the 1930s, there was evidence for the generality and replicability of a FFM of personality that consisted of surgency/extraversion (E), agreeableness (A), conscientiousness (C), neuroticism (N), and intellect/openness (O). Over a period of almost 50 years, a series of cumulative studies may be traced (e.g., Fiske, 1949; Norman, 1963; Tupes & Christal, 1961) that led to an emerging consensus on the utility and generalizability of this five-factor representation (e.g., Digman, 1979; Goldberg, 1977; Wiggins, 1973b). By the 1980s, there were three distinctive interpretations of the nature of these five factors (Costa & McCrae, 1985; Goldberg, 1981; Hogan, 1986), that were followed in the 1990s by still more interpretive perspectives (e.g., Buss, 1996; Wiggins & Trapnell, 1996). It should be noted, however, that although there are a number of workers within the multivariate paradigm who agree on the number, if not the nature, of the factors in the FFM, there are many who strongly contest both the number of factors and their significance (e.g., Block, 1995; Eysenck, 1992; Hough, 1992; Waller & Ben-Porath, 1987).

4.12.4.2 Conceptual Framework

Cattell's (1957) master plan for constructing a theory of personality presupposed a rigorous and representative sampling of the language personality sphere that, in effect, defined the universe of content of behaviors to be explained (Wiggins, 1984). To make certain that this universe of content was properly defined, Norman (1967) replicated the earlier lexical research of Allport and Odbert (1936), using a more recent unabridged dictionary, and subsequently employing more rigorous clustering procedures than those used by Cattell (see John, 1990). Norman's taxonomy of personality attributes was subsequently employed by Goldberg (1981, 1990, 1993) in a program of methodologically elegant research on the natural language of personality that firmly established what he termed the "Big Five" dimensions identified by earlier investigators.

Goldberg's taxonomic research program revitalized interest in the five-factor model and within a relatively brief period of time a variety of new instruments were developed (e.g., Botwin & Buss, 1989; Costa & McCrae, 1985; Goldberg, 1992; Hogan, 1986; Trapnell & Wiggins, 1990) that measured these five dimensions within an equal variety of conceptual frameworks, for example, evolutionary theory (Buss, 1996), trait theory (McCrae & Costa, 1996), lexical theory (Saucier & Goldberg, 1996), socioanalytic theory (Hogan, 1996), and interpersonal theory (Wiggins & Trapnell, 1996). Moreover, a series of investigations by Costa and McCrae revealed that instruments employed in the other four traditions of personality assessment—psychodynamic (Myers & McCaulley, 1985), personological (Jackson, 1984), empirical (Hathaway & McKinley, 1983), and interpersonal (Wiggins, Trapnell, & Phillips, 1988)—showed meaningful convergences with some or all of the dimensions of the FFM (Wiggins & Trapnell, 1997).

4.12.4.3 Assessment Instruments

The NEO Personality Inventory (NEO PI-R; Costa & McCrae, 1992b) is the instrument of choice for measuring the dimensions of the FFM of personality. This test provides global measures (domain scores) for each of the five factors and six more specific measures (facet scores) within each of the five domains: (i) neuroticism (anxiety, angry hostility, depression, self-consciousness, impulsiveness, vulnerability); (ii) extraversion (warmth, gregariousness, assertiveness, activity, excitement-seeking, positive emotions); (iii) openness (fantasy, aesthetics, feelings, actions, ideas, values); (iv) agreeableness (trust, straightforwardness, altruism, compliance, modesty, tender-mindedness); and (v) conscientiousness (competence, order, dutifulness, achievement striving, self-discipline). The "astonishingly fruitful research collaboration" (Block, 1995) of Costa and McCrae has produced scores of publications that document the substantive, structural, and empirical validity of this instrument.

4.12.4.4 Interpretive Principles

There are notable differences in interpretive principles suggested by proponents of different multivariate instruments and these would appear to reflect different claims as to the extensiveness and importance of the "universe of content" measured by each instrument, and the theoretical basis on which that content is interpreted for each instrument. With respect to the former, the differences among alternative theories of personality structure within the multivariate tradition have often reflected preferences for different methods of factoring correlations among the phenotypic variables of personality study. Thus, Cattell, who argued for 16 primary source traits, as well as the originator of the multiple-factor method himself (Thurstone, 1934), have both been accused of "overfactoring" (Digman, 1996); while Eysenck (1992), who disputes anything beyond three factors, has been accused of "underfactoring" (Costa & McCrae, 1992a). That almost half a century of essentially methodological disagreements may be approaching resolution can be seen in the emerging consensus that the FFM of personality is an appropriate "working model" of the universe of content of personality structure that may (and should) be expanded, as necessary, whenever the incremental validity of additional dimensions are demonstrated (Wiggins, 1996).

The theoretical bases on which a five-factor profile may be interpreted are of greater interest to the practitioner. And, as is true of a number of instruments that originated in other personality assessment traditions, it is possible to interpret the FFM from a variety of theoretical perspectives. The NEO PI-R (Costa & McCrae, 1992b) provides explicit principles for interpreting normal and abnormal behavior from the standpoint of modern trait theory (McCrae & Costa, 1996). The Hogan Personality Inventory (Hogan & Hogan, 1992) provides explicit interpretive principles for personnel selection in the workplace from the standpoint of socioanalytic theory (Hogan, 1996). A new edition of the venerable Sixteen Personality Factor Questionnaire (Cattell, Cattell, & Cattell, 1993) provides interpretive principles for five "global factor scales" (Cattell, 1994) from the standpoint of a long-standing "mainstream" personality theory (Wiggins, 1984); and the possibilities for interpreting other well-established instruments, such as the Personality Research Form (Jackson, 1984), from an FFM perspective, are under active investigation (e.g., Jackson, Paunonen, Fraboni, & Goffin, 1996).

4.12.4.5 Applications and Current Status

Historically, there has been little consensus regarding the relation between the normal dimensions of personality, measured by inventories developed within the multivariate tradition, and the dimensions and/or categories of psychopathology. Recently, however, there has been an increased conceptual and empirical focus on this issue, stimulated in part by the definition of personality disorders in terms of personality traits by the American Psychiatric Association (APA; 1980) and by the availability of new instruments and techniques for addressing this problem (Strack & Lorr, 1994).

It would appear that the global dispositions measured by the NEO PI-R provide a comprehensive assessment of emotional (N), interpersonal (E & A), motivational (C), and cognitive (O) styles that are not measured by traditional clinical instruments. As Costa (1991b) noted, "This portrait can be used in psychotherapy: to formulate a tentative diagnosis; to aid the clinician's development of empathy; to help select appropriate treatments; to identify the client's strengths; and to anticipate the outcome, duration, and course of therapy" (p. 395). A detailed and perceptive account of how this might be done has been provided by Miller (1991). The notable longitudinal stability of NEO PI-R dimensions in the adult personality (McCrae & Costa, 1990) has led to a revised view of the therapeutic enterprise:

> we expect that clients will bring these dispositions with them to psychotherapy, and that the therapist should (a) take them into account when trying to understand the individual and his or her problems, and (b) tailor the therapy to fit the needs and styles of the client. (McCrae, 1991, p. 406)

The most extensive psychodiagnostic application of the FFM in general, and of the NEO PI-R in particular, has been to the diagnosis and understanding of personality disorders (Costa & Widiger, 1994). The definition of these disorders in terms of dysfunctional personality traits, by a committee of the APA (1980), stimulated an unprecedented collaborative effort among psychiatrists, psychologists, and psychometricians to find common grounds for characterizing these disorders. It was found, for example, that both the APA diagnostic criteria and the clinical literature on personality disorders were compatible with descriptions of each disorder generated from the facet and domain scales of the NEO PI-R (see Widiger, Trull, Clarkin, Sanderson, & Costa, 1994, Table 1, p. 42).

Subsequent advances in relating the FFM to the personality disorders are too numerous to summarize, but may be found in current journals, particularly the *Journal of Personality Disorders*. The current literature may be viewed as reflecting yet another milestone in the long-standing relation between the FFM and the multivariate paradigm, which Digman (1994) characterized as follows:

> Now, after many years of lying on the closet shelf of personality theory, the model has been dusted off, "as good as new," and appears to be for many researchers ... a very meaningful theoretical structure for organizing the myriad specifics implied by the term *personality*. (p. 13)

4.12.5 EMPIRICAL TRADITION

4.12.5.1 Historical Background

The MMPI (Hathaway & McKinley, 1943) is, and has been for many years, the most widely used inventory of personality and psychopathology. To understand the 50-year history

of this instrument is to understand the history of objective personality assessment for the same time period. As Craik (1986) put it, in his historical survey of personality research methods, "the MMPI came to serve as the centerpiece of this period's [post-WWII] predominant or mainstream agenda" (p. 21). Within that agenda, one can discern the operation of two separate dialectical processes over time: ongoing disputes between the developers of the MMPI and their critics and bipolar shifts in conceptualization that have occurred within the evolving conceptual and interpretive frameworks of the developers of the MMPI. The ongoing disputes reflect the fact that, because of its prominence, the MMPI came to be associated with such contentious issues as clinical versus statistical prediction (Meehl, 1954) and response styles (Wiggins, 1962). The shifts in conceptualization reflect the virtuosity and intellectual flexibility of the developers of the MMPI who were able to shift from typological categories to continuous trait dimensions; from discriminant validity of differential diagnoses to the construct validity of scales and profiles; and from denigration of self-reports to the canonization of item content (Ben-Porath, 1994).

The dialectical processes just described were enacted by a remarkably diverse group of scientists/practitioners whose program of clinical research had its roots in the unique combination of behavioral, biological and psychometric thinking that prevailed at the University of Minnesota in the 1930s and 1940s (Meehl, 1989). Stark R. Hathaway was an originator of the MMPI, a mentor for successive generations of Minnesota students and a merciless critic of his own instrument (e.g., Hathaway, 1972). Paul E. Meehl—learning theorist, philosopher of science, psychoanalyst, and taxonometrician—is a legend in his own time, the principal theorist of the MMPI, and arguably the major figure in the field of personality assessment since the late 1950s. W. Grant Dahlstrom is the Talmudic scholar of the MMPI, whose meticulous organizations of that vast literature have informed and influenced generations of clinicians and researchers (e.g., Dahlstrom & Dahlstrom, 1980; Dahlstrom, Welsh, & Dahlstrom, 1972–1975; Welsh & Dahlstrom, 1956). James N. Butcher is a distinguished editor, scholar, and researcher who is principal author of the revised MMPI (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989), the new MMPI content scales (Butcher, Graham, Williams, & Ben-Porath, 1990), the adolescent version of the MMPI (MMPI-A; Butcher et al., 1992), and the MMPI automated clinical report (Butcher, 1993).

4.12.5.2 Conceptual Framework

The rationales underlying four distinguishable strategies of objective (as opposed to projective) test construction are based on differing views regarding the meaning of subjects' responses to objective test items (Wiggins, 1973b)

(i) *The rational (correspondence) strategy.* This assumes a one-to-one correspondence between a subject's "report" ("I am anxious") and a palpable internal state of the subject (the experience of anxiety) (Buchwald, 1961). This assumption was the cornerstone of introspectionism in early experimental psychology and it was the principal rationale underlying the earliest personality inventories (e.g., Woodworth, 1917).

(ii) *The empirical (instrumental) strategy.* This views a subject's "utterance" as "an intrinsically interesting and significant bit of verbal behavior, *the nontest correlates of which must be discovered by empirical means*" (Meehl, 1945, p. 297, emphasis added). If it were established empirically that a false response to the item "I am anxious" was more characteristic of hospitalized hysterics than of a normal control group, that particular bit of verbal behavior would become meaningful, and would illustrate the instrumental value of verbal behavior as a tool in psychodiagnostics. This philosophy was evident in the early behavioristic critique of introspectionism in experimental psychology and in later critiques of existing rational personality inventories of the 1930s (Humm & Wadsworth, 1935; Landis & Katz, 1934; Landis, Zubin, & Katz, 1935). The earliest example of a major inventory constructed under an empirical strategy was the Vocational Interest Blank (Strong, 1927), which served as the prototype for the construction of the MMPI.

(iii) *The constructive (substantive) strategy.* This views a subject's response as a manifestation of an underlying personality construct that is embedded in an interlocking system of laws that relate constructs to one another and to observable properties of the environment (Cronbach & Meehl, 1955). The meaning of a construct is given by the empirical laws into which it enters and the significance of a construct is given by the number of such lawful relations discovered. The construct validity of a personality scale is estimated from "the proportion of test score variance that is attributable to the construct variable" (p. 289).

(iv) *The interpersonal (self-presentational) strategy.* This views a subject's response as an interpersonal communication between the subject and the tester (or the institution that the tester represents) (Carson, 1969b; Leary, 1957). The meaningfulness of such a communication is dependent upon both the subject's and the tester's views of the meaning of test responses in a particular testing situation (e.g., hospital admission, vocational guidance). Thus, the subject may mean to communicate a complaint ("I am anxious") and the tester (or the scoring system employed by the institution) may encode this as "neuroticism." The subject may view the testing situation as an opportunity for self-disclosure (Jourard, 1964) and the examiner may view it as an opportunity for impression management (Goffman, 1959). Elsewhere it has been argued that the interpersonal view of test responses has much to commend it, not the least of which is the likelihood that this is the frame of reference which subjects themselves adopt (Wiggins, 1966).

4.12.5.3 Assessment Instruments

Construction of the MMPI began with the generation of a large pool of items that were considered representative of "behaviors of significance to the psychiatrist" (Hathaway & McKinley, 1940). These items were presented in true–false format to several groups of diagnosed psychiatric inpatients (e.g., depressives) and to a large group of normal control subjects. The proportion of patients in a given diagnostic group who responded "true" ("endorsement") to each item (e.g., 80%) was compared with the proportion of normal subjects who had responded "true" to that item (e.g., 10%), in order to identify discriminating items that might be keyed (in this instance, "true") on a clinical scale (e.g., depression scale; Hathaway & McKinley, 1942). Innovative procedures were then applied in the development of three validity scales that served to identify sources of invalidity or "faking" in test responding: a "lie scale" (*L*) with items of high desirability and low endorsement, an "infrequency scale" (*F*) with items of low desirability and low endorsement, and a subtle "correction scale" (*K*) that measures defensiveness when high and self-criticalness when low. The resultant three validity scales, eight clinical scales (hypochondriasis, depression, hysteria, psychopathic deviate, paranoia, psychasthenia, schizophrenia, hypomania), a masculinity/femininity scale, and a social introversion scale constitute the MMPI profile that was normed and standardized on the original sample of normal control subjects.

4.12.5.4 Interpretive Principles

4.12.5.4.1 Empirical strategy

The original purpose of the MMPI was to aid psychiatrists in their assignment of patients to the traditional psychiatric diagnostic categories of that time. The conservative rationale of the empirical strategy permitted such statements as "This subject's pattern of verbal behavior resembles more closely that of a group of hospitalized schizophrenics than it does that of a normal control group." This fact, together with evidence from other sources, was suggestive of the working diagnosis of "schizophrenia." Unfortunately, the individual clinical scales were not successful in subsequent attempts to discriminate diagnostic groups from normals (Hathaway, 1960), due most likely to lack of power in the original statistical comparisons and to unreliability of both predictor and criterion measures.

4.12.5.4.2 Constructive strategy

Ben-Porath (1994) provides an excellent historical summary of the "reinvention of the MMPI" following the failure of the empirical strategy and of its subsequent "evolution into an omnibus measure of personality" within a construct-oriented perspective. The highlights of this shift over a 50-year period of unprecedented research and clinical work include emphases upon: (i) MMPI profile configurations, rather than on single scales (e.g., Gough 1946); (ii) normal personality correlates of profile types (e.g., Black, 1956); (iii) additional empirically and substantively derived scales that broadened the nomological network of MMPI investigation (e.g., Morey, Waugh, & Blashfield, 1985), and (iv) actuarial prediction systems based on configural profile types (e.g., Marks & Seeman, 1963).

4.12.5.4.3 Interpersonal strategy

Loevinger's (1957) formulation of the substantive component of construct validity provided an even greater contrast to the original empirical perspective of the MMPI. Nevertheless, workers within the MMPI tradition for the most part continued to pursue the other components of construct validity just described. Interest in the "content" of patients' communication of complaints via the MMPI remained minimal until the appearance of Wiggins' (1966) MMPI Content Scales. In terms of Loevinger's distinctions, these content scales appeared to provide: (i) representative coverage of the universe of content of the MMPI item pool (see Johnson, Butcher, Null, & Johnson, 1984), (ii) psychometrically sound measures of 13 dimensions of self-report that were interpretable with reference to previous studies of the factorial structure of the MMPI (Welsh, 1956), and (iii) empirical evidence of convergent and

discriminant validity with reference to psychiatric diagnostic categories (Wiggins, 1966; Payne & Wiggins, 1972). Perhaps of greater interest to the MMPI community was the potential usefulness of these content scales in individual psychodiagnostic appraisal (e.g., Nichols, 1987). Variants of the original content scales are now an official supplement to the clinical scales of MMPI-2 (Butcher et al., 1990) and their incremental contribution to the differential diagnosis of psychopathology appears highly promising (e.g., Ben-Porath, Butcher, & Graham, 1991).

4.12.5.5 Applications and Current Status

The MMPI developed in the 1940s was, by contemporary standards, based on inadequate and outdated norms, replete with offensive and anachronistic items, and patently unsuited to the diagnostic task for which it was devised. The clinical scales were derived by less than optimal contrasted group procedures (Jackson, 1971) that, from the beginning, appeared unreplicable (Benton, 1949). In the absence of a coherent theoretical rationale, the profiles, code-types, and actuarial systems based on these scales have an ambiguous conceptual status (Helmes & Reddon, 1993). Pressures to restandardize, revise, or replace this relic have existed almost from its inception, culminating in a historically important summit meeting (Butcher, 1972) in which the third option was seriously considered (Wiggis, 1973a). Once again, however, the resilient and perdurable Minneapolis Group (Butcher et al., 1989) rose to the challenges from within and without to produce what was optimistically dubbed "MMPI-2" (leaving open the possibility of MMPI-3 in the year 2035).

Although the original MMPI is a disgruntled psychometric critic's dream come true, it is difficult to ignore the past and present status of the MMPI as our most widely used objective personality test. Given what appeared to be the lesser of two evils, the personnel of the MMPI Restandardization Project opted to emphasize continuity with the vast interpretive and empirical database accumulated for the original clinical and validity scales while simultaneously removing inappropriate items, correcting psychometric deficiencies where possible, and generating new content scales. Shorn of its dated and politically incorrect items, the lamblike content of the MMPI-2 is now clothed in the lupine garments of computer technology and state-of-the-art methodology (e.g., uniform *T*-scores, adaptive testing, measurement of variable response inconsistency). MMPI-2 appears to be replacing the original in most clinical settings (e.g., Webb, Levitt, & Rojdev, 1993), although there continue to be criticisms from both without and within the MMPI community (e.g., Caldwell, 1997).

In 1972, Hathaway concluded: "So, in summary, we are stuck, I think, with the MMPI ... for a dreary while longer, although a prophet may even now be wandering in the psychometric wilderness" (1972 p. 40). There is little doubt that we will be "stuck" with MMPI-2 for many years to come, although the paths of the prophets are now more discernible (e.g., Costa, 1991a; Morey, 1991).

4.12.6 INTERPERSONAL PARADIGM

4.12.6.1 Historical Background

The interpersonal circumplex model had its origins in the writings of Harry Stack Sullivan (1953) who introduced a number of original and radical ideas into the American psychiatry of the 1940s. Although not entirely liberated from Freud's spatial-hydraulic metapsychology, Sullivan was "behavioral" in the sense that his primary emphasis was upon the things that persons do to one another in interpersonal transactions, rather than upon internal processes that persons might "have." His oft-cited definition of personality as "the relatively enduring pattern of recurrent interpersonal situations which characterize a human life" represents an even greater departure from psychoanalytic thought, as well as a unique perspective within personality theory itself.

Sullivan was influenced by Lewin's (1938) field-theoretical conceptualization of contemporaneous, bidirectional influences in "psychological fields," which led him to define the interpersonal situation as the basic unit of observation in psychiatry. This idea does not connote a separate "person" or "persons" who may be considered independently from the complex field (situation) of bidirectional causalities in which they are embedded; nor does it hold out the hope of capturing such complexities as "interaction" terms in analyses of variance.

The Kaiser Foundation Research Project, in which Timothy Leary (1957) was the most preeminent investigator, attempted to operationalize Sullivanian concepts in terms of concrete measurement procedures. An ordinary language analysis of clinicians' observations of the things that patients did to each other (and to themselves) in group psychotherapy led to a taxonomy of interpersonal behaviors that appeared to be empirically well-captured by a circular arrangement of interpersonal variables organized around the coordinates of

"dominance" and "affiliation." LaForge, Leary, Naboisek, Coffey, and Freedman (1954) derived the basic trigonometry of what would later be called a "circumplex," without knowledge of Guttman's (1954) related work on that topic.

Notable contributions during the 1960s included the conceptualization of both maternal and child behavior within an interpersonal circumplex framework (Schaefer, 1961), a psychometrically sophisticated replication of the circumplex within a clinical population (Lorr & McNair, 1963), and a highly influential integration of the circumplex with the clinical, social, and experimental psychology of that time (Carson, 1969a). Alternative conceptual formulations of the interpersonal model were presented in the 1970s (e.g., Benjamin, 1974; Kiesler, 1979; Wiggins, 1979), and during the 1980s the model was applied to psychotherapy (Anchin & Kiesler, 1982), complementarity (Kiesler, 1983), and interpersonal problems (Horowitz, Rosenberg, Baer, Ureno, & Villasenor, 1988). Among the contributions of the 1990s have been an updated and psychometrically-sound version of the original interpersonal checklist (Wiggins, 1995), a comprehensive exposition of contemporary interpersonal theory and research (Kiesler, 1996), and a presentation of the impressive variety of contexts in which the interpersonal circumplex model has been applied (Plutchik & Conte, 1997).

4.12.6.2 Conceptual Framework

The radical interpersonalism of Sullivan (1953) may be difficult to comprehend on first exposure because of our ingrained "individualist language" (p. 50) for describing personality as reflecting attributes of a discrete individual who is "separate" from others and from a shared social environment. Theorists within the interpersonal paradigm have attempted to operationalize Sullivan's conceptualization of personality in ways that avoid this individualistic bias, by defining personality as "nothing more (or less) than the patterned regularities that may be observed in an individual's relations with other persons, who may be real in the sense of actually being present, real but absent and hence 'personified' or 'illusory'" (Carson, 1969a, p. 26).

On a metatheoretical level, it is helpful to think of the dominance and affiliation axes of the interpersonal circumplex in terms of Bakan's (1966) concepts of "agency" (mastery, self-assertion, self-expansion) and "communion" (union with a larger social or spiritual entity), respectively (Wiggins, 1991). These concepts are compatible with Sullivan's emphasis on the communal context of interpersonal transactions and with his belief in the importance of the social sciences for understanding interpersonal situations. When conjoined with Foa and Foa's (1974) theory of the exchange of love (communion) and status (agency) in interpersonal transactions, these concepts provide a rich conceptual framework for the measurement and understanding of interpersonal behavior (Wiggins & Trapnell, 1996).

4.12.6.3 Assessment Instruments

4.12.6.3.1 Interpersonal Adjective Scales

The Interpersonal Ajective Scales (IAS; Wiggins, 1995) evolved from a psychological taxonomy of trait-descriptive terms that was developed within the framework of a larger program of collaborative research on language and personality (Goldberg, 1977). Within a representative pool of trait-descriptive adjectives selected from an unabridged dictionary, an "interpersonal domain" was distinguished from other domains, such as characterological, temperamental, and cognitive domains (Wiggins, 1979). Approximately 800 interpersonal adjectives were assigned to the original categories of the interpersonal circumplex (Leary, 1957) on both conceptual and empirical grounds. Using computer-based multivariate procedures it was found that scales based on the original categories failed to meet certain circumplex criteria that could be better met with scales based on the revised categories that now constitute the IAS (Wiggins, 1995). The IAS consists of 64 adjectives that respondents rate for self-descriptive accuracy on an eight-place Likert scale ranging from "extremely inaccurate" to "extremely accurate." Eight scales, or "octants," of eight items each, assess the interpersonal dispositions listed in the first column of Table 1. The PA octant, for example, includes items such as "dominant," "forceful," and "assertive."

4.12.6.3.2 Inventory of Interpersonal Problems

Horowitz (1979) transcribed statements of interpersonal problems expressed by psychiatric outpatients in the course of videotaped intake interviews and employed these statements as items in the construction of the Inventory of Interpersonal Problems (IIP), in which respondents are required to indicate the extent to which each of 127 statements is problematic on a five-place Likert scale ranging from "not at

Table 1 Octant scales from three interpersonal assessment instruments.

	IAS Interpersonal Adjective Scales	*IAS Inventory of Interpersonal Problems*	*IAS Impact Message Inventory*
PA	Assured–dominant	Domineering	Dominant
BC	Arrogant–calculating	Vindictive	Hostile–dominant
DE	Cold-hearted	Cold	Hostile
FG	Aloof-introverted	Socially avoidant	Hostile–submissive
HI	Unassured–submissive	Nonassertive	Submissive
JK	Unassuming–ingenuous	Exploitable	Friendly–submissive
LM	Warm–agreeable	Overly nurturant	Friendly
NO	Gregarious–extraverted	Intrusive	Friendly–dominant

all" to "extremely" (Horowitz et al., 1988). Subsequently, Alden, Wiggins and Pincus (1990) developed a circumplex version of this inventory (IIP-C) that consists of eight scales, of eight items each, that assess the interpersonal problems listed in the second column of Table 1. The PA octant, for example, includes items such as "I try to control other people too much."

4.12.6.3.3 Impact Message Inventory

The Impact Message Inventory (IMI) is a highly original and promising method of assessment that is based on Kiesler's (1988) theory of interpersonal communication in psychotherapy. The theory postulates that disordered individuals are unaware of the unintended, inappropriate, and ambiguous messages they repetitively "send" to others and that they are therefore confused and distressed by the pattern of negative responses they consistently evoke or "pull" from others. The IMI attempts to identify the location within the interpersonal circumplex of these patterns of negative response evoked in others, as a means of gaining insight into a client's maladaptive transactional behavior.

Items were generated from a content analysis of free responses to 15 interpersonal vignettes that described characters enacting 15 different interpersonal styles, similar to those found in the first column of Table 1. Respondents were asked to imagine themselves in the company of each of these characters and to record their covert reactions using the stem, "He makes me feel . . ." Content analysis of responses suggested three categories of covert reaction: direct feelings, action tendencies, and perceived evoking messages. The Octant version of the IMI consists of six items for each of the target stimuli listed in the third column of Table 1. For each octant, there are two items for each of the three categories of covert reaction to the target stimulus. Thus, for example, the PA scale includes two items each for direct feelings (e.g., "bossed around"), action tendencies (e.g., "I want to tell him to give someone else a chance to make a decision"), and perceived evoking messages (e.g., "He thinks he's always in control of things"). Respondents (e.g., a psychotherapist) are asked to imagine themselves in the company of a particular person (e.g., a psychotherapy patient) and to indicate the extent to which they experience the covert reactions on a four-place scale. Covert reactions to a target person (e.g., feeling "bossed around") are thus revealing of the personality and the interpersonal impact of that target person (e.g., rigidly dominant, PA).

4.12.6.4 Interpretive Principles

One of the most valuable features of the circumplex model is the opportunity it provides for interpreting interpersonal variables with reference to the geometric principles of a circle (LaForge et al., 1954). The variables of traditional multiscale inventories are typically displayed as "factor lists" (Hogan, 1983) in which the ordering and interrelations among variables lack both conceptual and interpretive significance. In contrast, the trigonometric procedures that can be applied to the variables of an interpersonal circumplex permit descriptions and diagnostic inferences that cannot be generated from traditional scales (Wiggins, Phillips, & Trapnell, 1989).

Some general principles of circumplex interpretation may be illustrated with reference to Figure 1, which presents the IAS profile of a 44-year-old woman who was employed as a senior bank manager. At the bottom of the figure are *T*-scores on the octant variables, expressed with reference to an appropriate normative group. These scores have been plotted on shaded sectors of the circle and they are interpreted as representing eight vectors in two-dimensional space. The mean or average directionality of these eight vectors is of critical diagnostic significance because it determines

the "typological" category to which this woman will be assigned. By trigonometric procedures, this average directionality was determined to be 90°, which falls exactly at the midpoint of the assured-dominant category (PA). Such an angular location is considered to be prototypical of individuals classified as pure "PA types." Had the location been at 110°, a more hostile manifestation of dominance (BC) would have been suggested; had it been at 65°, a more affiliative expression of dominance (NO) would have been suggested. But the protypicality of the present classification increases our confidence in asserting that the principal interpersonal style of this woman is one that emphasizes the exercise of power over others in a social context, by such activities as taking charge, making decisions, and winning arguments (Wiggins, 1995, p. 22).

The general shape of the profile in Figure 1 closely resembles the characteristic configuration of IAS profiles that is found in the average profiles of subjects in all IAS typological groups (Wiggins et al., 1989). The characteristic configuration of an IAS profiles is one in which the principal elevation occurs on the defining octant (in this case, PA), with secondary and approximately equal elevations occurring on adjacent octants (BC and NO), and diminishing and approximately equal elevations occurring on subsequent pairs of octants (DE and LM), (FG and JK), down to a highly truncated "opposite" (to the principal) octant (HI). Thus, across interpersonal situations, we would expect this woman frequently to behave in a forceful, assertive, dominant, and self-confident manner (PA); to somewhat less frequently behave in aggressive (BC) and gregarious (NO) ways; to seldom, if ever, behave in a submissive (HI) fashion, and so forth.

The vector length of an interpersonal profile is a measure of "deviance" in both a statistical and a psychiatric sense. In the former sense, vector length is the standard deviation of the eight interpersonal variables that indicates the "intensity" with which a pattern of interpersonal behavior is expressed. In the latter sense, high-variance profiles are associated with interpersonal "rigidity" and are found most often in psychiatric groups. The length of the arrow indicating angular location in Figure 1 is the vector length of the profile, which is approximately 1.8 standard deviations above that of the normative group. This would suggest that the pattern of interpersonal behaviors described above would be expressed vigorously, and quite possibly rigidly or intemperately.

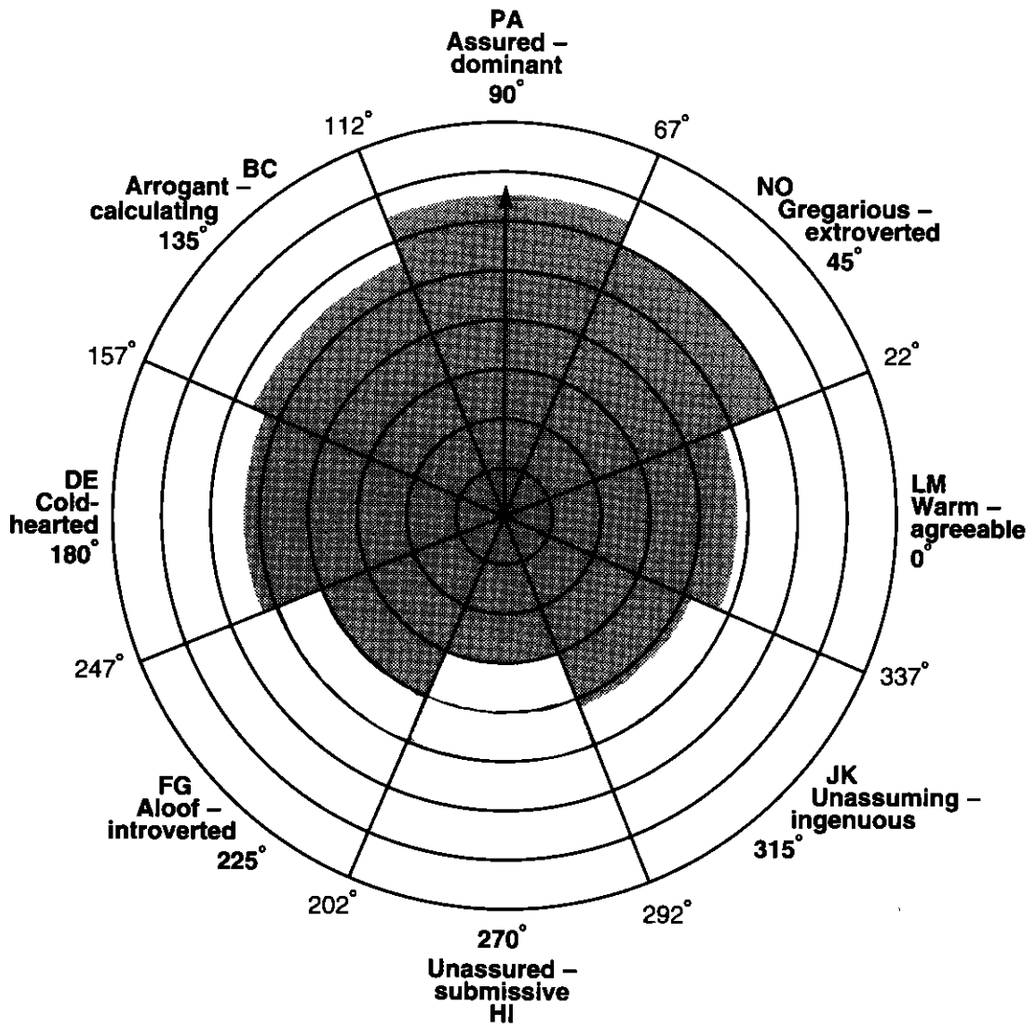
Having established that this individual's profile is representative of those obtained by assured dominant "types," and that this pattern of behavior is expressed in a clearly differentiated fashion, interpretation would proceed with reference to the empirical literature of the interpersonal tradition that has examined both the dimension of assured-dominant (PA) and the characteristics of dominant "types" in both normal and psychiatric populations (Kiesler, 1996). This would include the literature of other instruments that have studied this location (PA) in contexts such as interpersonal problems (IIP-C), impact messages (IMI), and other contexts that have been studied by interpersonal researchers.

4.12.6.5 Applications and Current Status

The definition of the personality disorders in terms of personality traits by the APA (1980), and the availability of new instruments and techniques for assessing personality disorder, has resulted in an increased understanding of this diagnostic axis. For example, many of the *Diagnostic and statistical manual of mental disorders* (*DSM*) personality disorder categories have been shown to be well captured by the two-dimensional structures of the IAS (e.g., Wiggins & Pincus, 1989, 1994) and the IIP-C (e.g., Pincus & Wiggins, 1990; Soldz, Budman, Demby, & Merry, 1993). Similarly, clinicians' ratings of *DSM* personality disorder categories have been found to be well captured by an interpersonal circumplex model (e.g., Blashfield, Sprock, Pinkston, & Hodgin, 1985; Plutchik & Conte, 1986). And more recently, Benjamin (1995) has provided a detailed and perceptive description of procedures for the diagnosis of personality disorders within the framework of her variant of the interpersonal circumplex model.

The interpersonal tradition originated in the context of psychotherapy and, perhaps more than any other tradition, has contributed to an understanding of the therapeutic process itself (e.g., Anchin & Kiesler, 1982; Benjamin, 1995; Kiesler, 1988; Safran & Segal, 1990; Sullivan, 1953). The most notable contribution of this tradition to psychotherapy has been the circumplex structural model that provides a framework for representing dyadic interactions in the therapeutic relationship. A number of different assessment devices based on the interpersonal circumplex, and variants of that model (e.g., Benjamin, 1974), have proven useful in studies of psychotherapy process and outcome (Kiesler, 1996).

Gurtman (1996) has emphasized the construct validity of the circumplex version of Horowitz's Inventory of Interpersonal Problems (IIP-C; Alden et al., 1990) within the psychotherapy context. Henry (1996) has made



	PA	BC	DE	FG	HI	JK	LM	NO
T-Score	65	56	52	40	30	41	49	60

Angular location = 90°
Vector length T-score = 68°

Figure 1 IAS profile of a 44-year-old bank manager.

a similar case for the Structural Analysis of Social Behavior (Benjamin, 1974) as a common metric for programmatic psychotherapy research. Kiesler, Schmidt, and Wagner (1997) have stressed the conceptual advantages of the IMI in analyzing the psychotherapeutic relationship. And Kiesler (1996) has also emphasized the potential of the revised Check List of Psychotherapy Transactions (Kiesler, Goldston, & Schmidt, 1991) for measuring the interpersonal behavior of interactants in the therapy relationship. Taken together, the studies summarized in the aforementioned review articles constitute an impressive empirical literature that attests to the heuristic potential of the interpersonal circumplex model in psychotherapy research. The future of the interpersonal circumplex tradition for research in both psychodiagnostics and psychotherapy appears to be a bright one (Wiggins & Trobst, 1997).

4.12.7 CONCLUSION

In the introduction to this chapter, we considered the diagnostic work of a hypothetical clinician who was both well trained and well informed about personality tests, but who was perhaps unaware of the conceptual and interpretive richness, and particularly of the distinctiveness, of the five personality assessment traditions just considered. Assuming that she persevered in reading through this highly condensed overview, there are a few "take home" messages that might now be conveyed to her.

First, in Kelly's (1955) terms, the five traditions arose and were developed within different "foci of convenience" and therefore are often most usefully applied within their specific areas of focus. The psychodynamic tradition of Freud arose from the revolutionary idea that the causes and reasons for behavior are frequently not apparent and that we must look "within" the individual, as it were, to discover the sources (often historical) of contemporary problems of living. The personological tradition arose in response to the question of "How shall a psychological life history be written?" The multivariate tradition arose as an extension of the mulifactor assessment of intelligence to the personality domain, and was subsequently concerned with the determination of the basic descriptive dimensions of personality. The empirical tradition arose in the context of descriptive psychiatry, and was originally most concerned with the differential diagnosis of psychopathology. The interpersonal tradition was originally focused on the character and quality of interpersonal relationships and on what persons did to one another within such relationships.

Second, instruments developed or employed within the five traditions have been found to have different "ranges of convenience" (Kelly, 1955), that is, to be generally or specifically applicable to a particular spectrum of assessment problems. Here it must be reiterated that there is not a one-to-one relation between instruments and traditions in this respect. A few examples must suffice: The Rorschach test may be useful in assessing cognitive and perceptual deficits; the WAIS may be useful in detecting covert motivational conflicts; the MMPI item pool has proven a fruitful source for the construction of literally hundreds of scales serving different assessment purposes; the TAT, and its associated need-press framework, has proven useful within many different traditions, as has the NEO PI-R, and the IAS.

Finally, our hypothetical clinician should be aware of the armamentarium of conceptual orientations and instruments that may be brought to bear on a specific referral or assessment task. It is extremely unlikely that there is one test, or one orientation, that will be optimal for all assessment tasks. Practical limitations, especially managed care in the USA, often preclude the once-favored idea of a "standard battery" of tests. It would also not be fair to expect our hypothetical clinician to master all available orientations and tests. But perhaps she could become familiar with a few more. Attending a workshop or two (outside of her present area of expertise) might enhance the professional life of our hypothetical clinician.

ACKNOWLEDGMENTS

We would like to express our indebtedness to Yossef S. Ben-Porath, Sidney J. Blatt, Dan P. McAdams, David S. Nichols, and William McKinley Runyan for their helpful suggestions concerning earlier drafts of sections within their own areas of expertise. However, we accept full responsibility for any errors of fact or interpretation that remain in the present version.

4.12.8 REFERENCES

Alden, L. E., Wiggins, J. S., & Pincus, A. L. (1990). Construction of circumplex scales for the Inventory of Interpersonal Problems. *Journal of Personality Assessment, 53,* 521–536.

Alexander, I. E. (1990). *Personology: Method and content in personality assessment and psychobiography.* Durham, NC: Duke University Press.

Allison, J., Blatt, S. J., & Zimet, C. N. (1968). *The interpretation of psychological tests.* New York: Harper and Row (Reprinted in 1988 by Taylor & Francis).

Allport, G. W. (1937). *Personality: A psychological interpretation.* New York: Holt, Rinehart, & Winston.

Allport, G. W. (1942). *The use of personal documents in psychological science.* New York: Social Science Research Council.

Allport, G. W. (1965). *Letters from Jenny.* New York: Harcourt, Brace, & World.

Allport, G. W. (1967). Autobiography. In E. G. Boring & G. Lindzey (Eds.), *A history of psychology in autobiography* (Vol. 5, pp. 1–25). New York: Appleton-Century-Crofts.

Allport, G. W., & Odbert, H. S. (1936). Trait names: A psycho-lexical study. *Psychological Monographs, 47* (Whole Issue No. 211).

American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

Anchin, J. C., & Kiesler, D. J. (Eds.) (1982). *Handbook of interpersonal psychotherapy.* Elmsford, NY: Pergamon.

Babcock, H. (1933). *A short form of the Babcock examination for the measurement of mental deterioration.* Chicago: Stoelting.

Bakan, D. (1966). *The duality of human existence: Isolation and communion in Western man.* Boston: Beacon Press.

Behrends, R. S., & Blatt, S. J. (1985). Internalization and psychological development through the life cycle. *Psychoanalytic Study of the Child, 40,* 11–39.

Benjamin, L. S. (1974). Structural analysis of social behavior. *Psychological Review, 81,* 392–425.

Benjamin, L. S. (1995). *Interpersonal diagnosis and treatment of personality disorders* (2nd ed.). New York: Guilford Press.

Ben-Porath, Y. S. (1994). The MMPI and MMPI-2: Fifty

years of differentiating normal and abnormal personality. In S. Strack & M. Lorr (Eds.), *Differentiating normal and abnormal personality* (pp. 361–401). New York: Springer.

Ben-Porath, Y. S., Butcher, J. N., & Graham, J. R. (1991). Contribution of the MMPI-2 content scales to the differential diagnosis of psychopathology. *Psychological Assessment, 3,* 634–640.

Benton, A. L. (1949). Review of Minnesota Multiphasic Personality Inventory. In O. K. Buros (Ed.), *The third mental measurements yearbook* (pp. 104–107). Highland Park, NJ: Gryphon Press.

Black, J. D. (1956). Adjectives associated with various MMPI codes. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine* (pp. 151–172). Minneapolis, MN: University of Minnesota Press.

Blashfield, R., Sprock, J., Pinkston, K., & Hodgin, J. (1985). Exemplar prototypes of personality disorder diagnoses. *Comprehensive Psychiatry, 26,* 11–21.

Blatt, S. J., & Allison, J. (1981). The intelligence test in personality assessment. In A. I. Rabin (Ed.), *Assessment with projective techniques* (pp. 137–231). New York: Springer.

Blatt, S. J., Brenneis, C. B. Schimek, J. G., & Glick, M. (1976). The normal development and psychopathological impairment of the concept of the object on Rorschach. *Journal of Abnormal Psychology, 85,* 364–373.

Blatt, S. J., & Lerner, H. D. (1983). The psychological assessment of object representation. *Journal of Personality Assessment, 47,* 7–28.

Block, J. (1995). A contrarian view of the five-factor approach to personality *Psychological Bulletin, 117,* 187–215.

Botwin, M. D., & Buss, D. M. (1989). Structure of act-report data: Is the five-factor model of personality recaptured? *Journal of Personality and Social Psychology, 56,* 988–1001.

Bowlby, J. (1973). *Attachment and loss, Vol. 2: Separation.* New York: Basic Books.

Buchwald, A. M. (1961). Verbal utterances as data. In H. Feigl & G. Maxwell (Eds.), *Current issues in the philosophy of science* (pp. 461–468). New York: Holt.

Buss, D. M. (1996). Social adaptation and five major factors of personality. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 180–207). New York: Guilford Press.

Butcher, J. N. (Ed.) (1972). *Objective personality assessment: Changing perspectives.* New York: Academic Press.

Butcher, J. N. (1993). *User's guide for the Minnesota Clinical Report.* Minneapolis, MN: National Computer Systems.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *The Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring.* Minneapolis, MN: University of Minnesota Press.

Butcher, J. N., Graham, J. R., Williams, C. L., & Ben-Porath, Y. S. (1990). *Development and use of the MMPI-2 content scales.* Minneapolis, MN: University of Minnesota Press.

Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., & Kaemmer, B. (1992). *Minnesota Multiphasic Personality Inventory (MMPI-A): Manual for administration, scoring, and interpretation.* Minneapolis, MN: University of Minnesota Press.

Caldwell, A. B. (1997). Whither goest our redoubtable mentor, the MMPI/MMPI-2? *Journal of Personality Assessment, 68,* 47–68.

Carson, R. C. (1969a). *Interaction concepts of personality.* Chicago: Aldine.

Carson, R. C. (1969b). Interpretive manual to the MMPI. In J. N. Butcher (Ed.), *MMPI: Research developments and clinical applications* (pp. 41–53). New York: McGraw-Hill.

Cattell, H. E. P. (1994). Development of the 16 PF fifth edition. In S. R. Conn & M. L. Rieke (Eds.), *The 16 PF fifth edition technical manual* (pp. 3–20). Champaign, IL: Institute of Personality and Ability Testing.

Cattell, R. B. (1933). Temperament tests. II: Tests. *British Journal of Psychology, 23,* 308–329.

Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. *Journal of Abnormal and Social Psychology, 38,* 476–506.

Cattell, R. B. (1949). *The Sixteen Personality Factor Questionnaire* (1st ed.). Champaign, IL: Institute of Personality and Ability Testing.

Cattell, R. B. (1957). *Personality and motivation structure and measurement.* Yonkers-on-Hudson, NY: World Book.

Cattell, R. B., Cattell, A., & Cattell, H. E. P. (1993). *Sixteen Personality Factor Questionnaire* (5th ed.). Champaign, IL: Institute of Personality and Ability Testing.

Costa, P. T., Jr. (Ed.) (1991a). Clinical use of the five-factor model of personality [Special series]. *Journal of Personality Assessment, 57*(3), 393–464

Costa, P. T., Jr. (1991b). Clinical use of the five-factor model: An introduction. *Journal of Personality Assessment, 57,* 393–398.

Costa, P. T., Jr., & McCrae, R. R. (1985). *The NEO Personality Inventory manual.* Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1992a). Four ways five factors are basic. *Personality and Individual Differences, 13,* 653–665.

Costa, P. T., Jr., & McCrae, R. R. (1992b). *NEO PI-R professional manual.* Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Jr., & Widiger, T. A. (Eds.) (1994). *Personality disorders and the five-factor model of personality.* Washington, DC: American Psychological Association.

Craik, K. H. (1986). Personality research methods: An historical perspective. *Journal of Personality, 54,* 18–51.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Dahlstrom, W. G., & Dahlstrom, L. E. (Eds.) (1980). *Basic readings on the MMPI: A new selection on personality measurement.* Minneapolis, MN: University of Minnesota Press.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972–1975). *An MMPI handbook.* (Vols. 1 & 2). Minneapolis, MN: University of Minnesota Press.

Digman, J. M. (1979, November). *The five major domains of personality variables: Analysis of questionnaire data in light of the five robust factors emerging from studies of rated characteristics.* Paper presented at the annual meeting of the Society of Multivariate Experimental Psychology, Los Angeles, CA.

Digman, J. M. (1994). Historical antecedents of the five-factor model. In P. T. Costa, Jr. & T. A. Widiger (Eds.), *Personality disorders and the five-factor model of personality* (pp. 13–18). Washington, DC: American Psychological Association.

Digman, J. M. (1996). The curious history of the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 1–20). New York: Guilford Press.

Elms, A. C. (1994). *Uncovering lives: The uneasy alliance of biography and psychology.* New York: Oxford University Press.

Erikson, E. H. (1950). *Childhood and society.* New York: Norton.

Erikson, E. H. (1958). *Young man Luther: A study in psychoanalysis and history.* New York: Norton.

Erikson, E. H. (1969). *Ghandi's truth: On the origins of militant nonviolence.* New York: Norton.

Eysenck, H. J. (1947). *Dimensions of personality.* London: Routledge & Kegan Paul.

Eysenck, H. J. (1953). *The structure of human personality.* New York: Wiley.

Eysenck, H. J. (1990). *Rebel with a cause.* London: Allen.

Eysenck, H. J. (1992). Four ways five-factors are *not* basic. *Personality and Individual Differences, 6,* 667–673.

Fairbain, W. R. D. (1952). *Psychoanalytic studies of the personality: The object relation theory of personality.* London: Routledge & Kegan Paul.

Fiske, D. W. (1949). Consistency of the factorial structure of personality ratings from different sources. *Journal of Abnormal and Social Psychology, 44,* 329–344.

Foa, U. G., & Foa, E. B. (1974). *Societal structures of the mind.* Springfield, IL: Charles C. Thomas.

Galton, F. (1888). Co-relations and their measurement. *Proceedings of the Royal Society, 45,* 135–145.

Gill, M. M. (Ed.) (1967). *The collected papers of David Rapaport.* New York: Basic Books.

Gill, M. M., & Klein, G. S. (1967). The structuring of drive and reality: David Rapaport's contributions to psychoanalysis and psychology. In M. M. Gill (Ed.), *The collected papers of David Rapaport* (pp. 8–34). New York: Basic Books.

Goffman, E. (1959). *The presentation of self in everyday life.* Garden City, NY: Doubleday Anchor.

Goldberg, L. R. (1977, August). *Language and personality: Developing a taxonomy of trait-descriptive terms.* Invited address to the Division of Evaluation and Measurement at the 86th annual meeting of the American Psychological Association, San Francisco.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 2, pp. 141–165). Beverly Hills, CA: Sage.

Goldberg, L. R. (1990). An alternative "Description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59,* 1216–1229.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4,* 26–34.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48,* 26–34.

Goldstein, K., & Scheerer, M. (1941). Abstract and concrete behavior: An experimental study with special tests. *Psychological Monographs, 53*(2) (Whole Issue No. 239).

Gough, H. G. (1946). Diagnostic patterns on the MMPI. *Journal of Clinical Psychology, 2,* 23–37.

Guilford, J. P. (1959). *Personality.* New York: McGraw-Hill.

Guilford, J. P., & Guilford, R. B. (1936). Personality factors S, E, and M, and their measurement. *Journal of Personality, 34,* 21–36.

Guilford, J. P., & Guilford, R. B. (1939). Personality factors D, R, T, and A. *Journal of Abnormal and Social Psychology, 34,* 21–36.

Gurtman, M. B. (1996). Interpersonal problems and the psychotherapy context: The construct validity of the Inventory of Interpersonal Problems. *Psychological Assessment, 8,* 241–255.

Guttman, L. (1954). A new approach to factor analysis: The radex. In P. R. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences* (pp. 258–348). Glencoe, IL: Free Press.

Hall, C. S., & Lindzey, G. (1978). *Theories of personality* (3rd ed.). New York: Wiley.

Hanfmann, E., & Kasanin, J. (1937). A method for the study of concept formation. *Journal of Psychology, 3,* 521–540.

Hartmann, H. (1939). *Ego psychology and the problem of adaptation.* New York: International Universities Press.

Hathaway, S. R. (1960). Forward. In W. G. Dahlstrom, G. S. Welsh, & L. E. Dahlstrom (Eds.), *An MMPI handbook: A guide to use in practice and research* (pp. vii–xi). Minneapolis, MN: University of Minnesota Press.

Hathaway, S. R. (1972). Where have we gone wrong? The mystery of the missing progress. In J. N. Butcher (Ed.), *Objective personality assessment: Changing perspectives* (pp. 21–43). New York: Academic Press.

Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I Construction of the schedule. *Journal of Psychology, 10,* 249–254.

Hathaway, S. R., & McKinley, J. C. (1942). A multiphasic personality schedule (Minnesota): III. The measurement of symptomatic depression. *Journal of Psychology, 14,* 73–84.

Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory.* Minneapolis, MN: University of Minnesota Press.

Hathaway, S. R., & McKinley, J. C. (1983). *The Minnesota Multiphasic Personality Inventory manual.* New York: Psychological Corporation.

Helmes, E., & Reddon, J. R. (1993). A perspective on developments in assessing psychopathology: A critical review of the MMPI and MMPI-2. *Psychological Bulletin, 113,* 453–471.

Henry, W. P. (1996). The structural analysis of social behavior as a common metric for programmatic psychotherapy research. *Journal of Consulting and Clinical Psychology, 64,* 1263–1275.

Hogan, R. (1983). A socioanalytic theory of personality. In M. M. Page (Ed.), *1982 Nebraska symposium on motivation: Personality—current theory and research* (pp. 55–89). Lincoln, NE: University of Nebraska Press.

Hogan, R. (1986). *Hogan Personality Inventory manual.* Minneapolis, MN: National Computer Systems.

Hogan, R. (1994). Reinventing ourselves (Review of D. P. McAdams, *The stories we live by*). *Contemporary Psychology, 39,* 355–356.

Hogan, R. (1996). A socioanalytic perspective on the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 163–179). New York: Guilford Press.

Hogan, R., & Hogan, J. (1992). *Hogan Personality Inventory manual.* Tulsa, OK: Hogan Assessment Systems.

Holt, R. R. (1951). The Thematic Apperception Test. In H. A. Anderson & G. L. Anderson (Eds.), *An introduction to projective techniques* (pp. 181–229). New York: Prentice-Hall.

Holt, R. R. (Ed.) (1967). Motives and thought: Psychoanalytic essays in honor of David Rapaport. *Psychological Issues, 5,* Monograph 18/19.

Holt, R. R., and Havel, J. (1960). A method for assessing primary and secondary process in Rorschach responses. In M. A. Rickers-Ovsiankina (Ed.), *Rorschach psychology* (pp. 263–315). New York: Wiley.

Horowitz, L. M. (1979). On the cognitive structure of interpersonal problems treated in psychotherapy. *Journal of Consulting and Clinical Psychology, 47,* 5–15.

Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G., & Villasenor, V. S. (1988). The Inventory of Interpersonal Problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56,* 885–892.

Hough, L. (1992). The "Big Five" personality variables—construct confusion: Description versus prediction. *Human Performance, 5,* 139–155.

Humm, D. G., & Wadsworth, G. W. (1935). The Humm-

Wadsworth temperament scale. *Journal of Applied Psychology, 92,* 163–200.
Jackson, D. N. (1971). The dynamics of structured personality tests: 1971. *Psychological Review, 78,* 229–248.
Jackson, D. N. (1984). *Personality Research Form manual* (3rd ed.). Fort Huron, MI: Research Psychologists Press.
Jackson, D. N., Paunonen, S. V., Fraboni, M., & Goffin, R. D. (1996). A five-factor versus six-factor model of personality structure. *Personality and Individual Differences, 20,* 33–45.
Jacobson, E. (1954). *The self and the object world.* New York: International Universities Press.
John, O. P. (1990). The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 66–100). New York: Guilford Press.
Johnson, J. H., Butcher, J. N., Null, C., & Johnson, K. N. (1984). Replicated item level factor analysis of the full MMPI. *Journal of Personality and Social Psychology, 47,* 105–114.
Jourard, S. M. (1964). *The transparent self.* Princeton, NJ: Van Nostrand.
Jung, C. G. (1971). *Psychological types* (H. G. Baynes, Trans.; revised by R. F. C. Hull). Princeton, NJ: Princeton University Press (Original work published 1923).
Kelly, G. A. (1955). *The psychology of personal constructs* (2 vols.). New York: Norton.
Kiesler, D. J. (1979). An interpersonal communication analysis of relationship in psychotherapy. *Psychiatry, 42,* 299–311.
Kiesler, D. J. (1983). The 1982 Interpersonal Circle: A taxonomy for complementarity in human transactions. *Psychological Review, 90,* 185–214.
Kiesler, D. J. (1988). *Therapeutic metacommunication: Therapist impact disclosure as feedback in psychotherapy.* Palo Alto, CA: Consulting Psychologists Press.
Kiesler, D. J. (1996). *Contemporary interpersonal theory and research: Personality, psychopathology, and psychotherapy.* New York: Wiley.
Kiesler, D. J., Goldston, C. S., & Schmidt, J. A. (1991). *Manual for the Check List of Interpersonal Transactions-Revised (CLOIT-R) and the Check List of Psychotherapy transactions-revised (CLOPT-R).* Richmond, VA: Virginia Commonwealth University.
Kiesler, D. J., & Schmidt, J. A. (1993). *The Impact Message Inventory: Form IIA Octant Scale version.* Palo Alto, CA: Mind Garden.
Kiesler, D. J., Schmidt, J. A., & Wagner, C. C. (1997). A circumplex inventory of impact messages: An operational bridge between emotion and interpersonal behavior. In R. Plutchik & H. R. Conte (Eds.), *Circumplex models of personality and emotions* (pp. 221–244). Washington, DC: American Psychological Association.
Kluckhohn, C., & Murray, H. A. (1953). Personality formation: The determinants. In C. Kluckhohn, H. A. Murray, & D. M. Schneider (Eds.), *Personality in nature, society, and culture* (pp. 53–67). New York: Knopf.
Kohut, H. (1971). *The analysis of the self: A systematic psychoanalytic approach to the treatment of narcissistic personality disorders.* New York: International Universities Press.
LaForge, R., Leary, T. F., Naboisek, H., Coffey, H. S., & Freedman, M. B. (1954). The interpersonal dimension of personality: II. An objective study of repression. *Journal of Personality, 23,* 129–153.
Landis, C., & Katz, S. E. (1934). The validity of certain questions which purport to measure neurotic tendencies. *Journal of Applied Psychology, 18,* 343–356.
Landis, C., Zubin, J., & Katz, S. E. (1935). Empirical validation of three personality adjustment inventories. *Journal of Educational Psychology, 26,* 321–330.
Leary, T. (1957). *Interpersonal diagnosis of personality.* New York: Ronald.
Lewin, K. (1938). *The conceptual representation and measurement of psychological forces.* Durham, NC: Duke University Press
Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3,* 635–694.
Lorr, M., & McNair, D. M. (1963). An interpersonal behavior circle. *Journal of Abnormal and Social Psychology, 67,* 68–75.
MacKinnon, D. W. (1944). The structure of personality. In J. McV. Hunt (Ed.), *Personality and the behavior disorders* (Vol. 1, pp. 3–48). New York: Ronald.
MacKinnon, D. W. (1975). IPAR's contribution to the conceptualization and study of creativity. In I. A. Taylor & J. W. Getzels (Eds.), *Perspectives in creativity* (pp. 60–89). Chicago: Aldine.
Main, M., Kaplan, N., & Cassidy, J. (1985). Security in infancy, childhood, and adulthood: A move to the level of representation. *Monographs of the Society for Research in Child Development, 50,* 66–104.
Marks, P. A., & Seeman, W. (1963). *The actuarial description of abnormal personality: An atlas for use with the MMPI.* Baltimore, MD: Williams & Wilkins.
Mayman, M. (1967). Object representations and object relationships in Rorschach responses. *Journal of Projective Techniques, 31,* 17–25.
McAdams, D. P. (1988). Biography, narrative, and lives: An introduction. *Journal of Personality, 56,* 1–18.
McAdams, D. P. (1989). *Intimacy: The need to be close.* New York: Doubleday.
McAdams, D. P. (1993). *The stories we live by: Personal myths and the making of the self.* New York: William Morrow.
McAdams, D. P. (1994). *The person: An introduction to personality psychology* (2nd ed.). Fort Worth, TX: Harcourt Brace.
McAdams, D. P. (1996). Personality, modernity, and the storied self: A contemporary framework for studying persons (Target article). *Psychological Inquiry, 7,* 295–321.
McClelland, D. C. (1961). *The achieving society.* New York: Van Nostrand.
McCrae, R. R. (1991). The five-factor model and its assessment in clinical settings. *Journal of Personality Assessment, 57,* 399–414.
McCrae, R. R., & Costa, P. T., Jr. (1990). *Personality in adulthood: Emerging lives, enduring dispositions.* New York: Guilford Press.
McCrae, R. R., & Costa, P. T., Jr. (1996). Toward a new generation of personality theories: Theoretical contexts for the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives.* (pp. 51–87). New York: Guilford Press.
Meehl, P. E. (1945). The dynamics of "structured" personality tests. *Journal of Clinical Psychology, 1,* 296–303.
Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and review of the evidence.* Minneapolis, MN: University of Minnesota Press.
Meehl, P. E. (1989). Autobiography. In G. Lindzey (Ed.), *A history of psychology in autobiography* (Vol. 3, pp. 337–389). Stanford, CA: Stanford University Press.
Miller, T. R. (1991). The psychotherapeutic utility of the five-factor model of personality: A clinician's experience. *Journal of Personality Assessment, 57,* 415–433.
Millon, T. (1969). *Modern psychopathology.* Philadelphia: Saunders.
Morey, L. C. (1991). *Personality Assessment Inventory manual.* Odessa, FL: Psychological Assessment Resources.
Morey, L. C., Waugh, M. H., & Blashfield, R. K. (1985). MMPI scales for *DSM-III* personality disorders: Their derivation and correlates. *Journal of Personality Assessment, 49,* 245–251.
Morgan, C. D., & Murray, H. A. (1935). A method for investigating fantasies: Thematic Apperception Test. *Archives of Neurology and Psychiatry, 34,* 209–306.
Murray, H. A. (1938). *Explorations in personality.* New York: Oxford University Press.
Murray, H. A. (1943). *The Thematic Apperception Test: Manual.* Cambridge, MA: Harvard University Press.
Murray, H. A. (1959). Preparations for the scaffold of a comprehensive system. In S. Koch (Ed.), *Psychology: A study of science* (Vol. 3, pp. 7–54). New York: McGraw-Hill.
Myers, I. B., & McCaulley, M. H. (1985). *Manual: A guide to the development and use of the Myers-Briggs Type Indicator.* Palo Alto, CA: Consulting Psychologists Press.
Nichols, D. S. (1987). Interpreting the Wiggins MMPI content scales. In K. L. Moreland & J. N. Butcher (Eds.), *Clinical notes on the MMPI* (No. 10, pp. 3–26). Minneapolis: National Computer Systems.
Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology, 66,* 574–583.
Norman, W. T. (1967). *2800 personality trait descriptors: Normative operating characteristics in a university population.* Ann Arbor, MI: University of Michigan, Department of Psychology.
Office of Strategic Services Assessment Staff (1948). *Assessment of men.* New York: Rinehart.
Payne, F. D., & Wiggins, J. S. (1972). MMPI profile types and the self-report of psychiatric patients. *Journal of Abnormal Psychology, 79,* 1–8.
Pearson, K. (1896). Mathematical contributions to the theory of evolution: Regression, heredity, and panmixia. *Philosophical Transactions, 187a,* 253–318.
Pincus, A. L., & Wiggins, J. S. (1990). Interpersonal problems and conceptions of personality disorders. *Journal of Personality Disorders, 4,* 342–352.
Plutchik, R., & Conte, H. R. (Eds.) (1997). *Circumplex models of personality and emotions.* Washington, DC: American Psychological Association.
Plutchik, R., & Conte, H. R. (1986). Quantitative assessment of personality disorders. In R. Michels & J. O. Cavenar, Jr. (Eds.), *Psychiatry* (Vol. 1, pp. 1–13). Philadelphia: Lippincott.
Prelinger, E., & Zimet, C. N. (1964). *An ego-psychological approach to character assessment.* Glencoe, IL: The Free Press.
Rapaport, D. (1942). Principles underlying projective techniques. *Character and Personality, 10,* 213–219 (reprinted in Gill, 1967).
Rapaport, D. (1944–1946). *Manual of diagnostic psychological testing* (2 vols.). New York: Josiah Macy, Jr. Foundation.
Rapaport, D. (1946). Principles underlying nonprojective tests of personality. *Annals of the New York Academy of Sciences, 46,* 643–652 (reprinted in Gill, 1967).
Rapaport, D. (1951). The conceptual model of psychoanalysis. *Journal of Personality, 20,* 56–81 (reprinted in Gill, 1967).
Rapaport, D. (1960). The structure of psychoanalytic theory: A systematizing attempt. *Psychological Issues, 2,* Monograph 6.
Rapaport, D., Gill, M., & Schafer, R. (1946). *Diagnostic psychological testing* (2 vols.). Chicago: Year Book.
Rapaport, D., Menninger, K. A., & Schafer, R. (1947). The new role of psychological testing in psychiatry. *American Journal of Psychiatry, 103,* 473–476 (reprinted in Gill, 1967).
Rorschach, H. (1921). *Psychodiagnostik.* Bern, Switzerland: Bircher. (Trans. Hans Huber Verlag, 1942).
Runyan, W. McK. (1982). *Life histories and psychobiography: Explorations in theory and method.* New York: Oxford University Press.
Safran, J. D., & Segal, Z. V. (1990). *Interpersonal process in cognitive therapy.* New York: Basic Books.
Sarbin, T. R. (Ed.) (1986). *Narrative psychology: The storied nature of human conduct.* New York: Oxford University Press.
Saucier, G., & Goldberg, L. R. (1996). The language of personality: Lexical perspectives on the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 21–50). New York: Guilford Press.
Schaefer, E. S. (1961). Converging conceptual models for maternal behavior and for child behavior. In J. G. Glidewell (Ed.), *Parental attitudes and child behavior* (pp. 124–146). Springfield, IL: Charles C. Thomas.
Schafer, R. (1948). *The clinical application of psychological tests.* New York: International Universities Press.
Schafer, R. (1954). *Psychoanalytic interpretation in Rorschach testing: Theory and application.* New York: Grune & Stratton.
Schafer, R. (1967). *Projective testing and psychoanalysis.* New York: International Universities Press.
Soldz, S., Budman, S., Demby, A., & Merry, J. (1993). Representation of personality disorders in circumplex and five-factor space: Explorations with a clinical sample. *Psychological Assessment, 5,* 53–63.
Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology, 15,* 201–293.
Stern, G. G., Stein, M. I., & Bloom, B. S. (1956). *Methods in personality assessment.* Glencoe, IL: Free Press.
Stewart, A. J., Franz, C., & Layton, L. (1988). The changing self: Using personal documents to study lives. *Journal of Personality, 56,* 41–74.
Strack, S. & Lorr, M. (1994). Introduction. In S. Strack & M. Lorr (Eds.), *Differentiating normal and abnormal personality* (pp. xiii–xviii). New York: Springer.
Strong, E. K., Jr. (1927). *Vocational Interest Blank (Form A).* Stanford, CA: Stanford University Press.
Sullivan, H. S. (1953). *The interpersonal theory of psychiatry.* New York: Norton.
Thurstone, L. L. (1934). The vectors of mind. *Psychological Review, 41,* 1–32.
Tomkins, S. S. (1947). *The Thematic Apperception Test.* New York: Grune & Stratton.
Tomkins, S. S. (1979). Script theory: Differential magnification of affects. In H. E. Howe, Jr. & R. A. Dienstbier (Eds.), *Nebraska symposium on motivation* (Vol. 26, pp. 201–236). Lincoln, NE: University of Nebraska Press.
Trapnell, P. D., & Wiggins, J. S. (1990). Extension of the Interpersonal Adjective Scales to include the Big Five dimensions of personality. *Journal of Personality and Social Psychology, 59,* 781–790.
Tupes, E. C., & Christal, R. E. (1961). *Recurrent personality factors based on trait ratings* (USAF ASD Tech. Rep. No. 61–97). Lackland Air Force Base, TX: US Air Force.
Vernon, P. E. (1950). *The structure of human abilities.* London: Methuen.
Waller, N. G., & Ben-Porath, Y. S. (1987). Is it time for clinical psychology to embrace the five-factor model of personality? *American Psychologist, 42,* 887–889.
Webb, J. T., Levitt, E. E., & Rojdev, R. (1993, March). *After three years: A comparison of the clinical use of the MMPI and MMPI-2.* Paper presented at the 53rd Annual Meeting of the Society for Personality Assessment, San Francisco, CA.
Wechsler, D. (1941). *The measurement of adult intelligence.* Baltimore: Williams & Wilkins.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore: Williams & Wilkins.
Welsh, G. S. (1956). Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine* (pp. 264–281). Minneapolis, MN: University of Minnesota Press.
Welsh, G. S., & Dahlstrom, W. G. (1956). *Basic readings on the MMPI in psychology and medicine*. Minneapolis, MN: University of Minnesota Press.
Westen, D. (1990). Psychoanalytic approaches to personality. In L. Pervin (Ed.), *Handbook of personality theory and research* (pp. 21–65). New York: Guilford Press.
Westen, D. (1991). Clinical assessment of object relations using the TAT. *Journal of Personality Assessment, 56*, 56–74.
White, R. W. (1952). *Lives in progress* (1st ed.). New York: Holt, Rinehart, & Winston.
White, R. W. (1966). *Lives in progress* (2nd ed.). New York: Holt, Rinehart, & Winston.
White, R. W. (1975). *Lives in progress* (3rd ed.). New York: Holt, Rinehart, & Winston.
White, R. W. (1981). Exploring personality the long way: The study of lives. In A. I. Rabin, J. Arnoff, A. M. Barclay, & R. A. Zucker (Eds.), *Further explorations in personality* (pp. 3–19). New York: Wiley.
Widiger, T. A., Trull, T. J., Clarkin, J. F., Sanderson, C., & Costa, P. T., Jr. (1994). A description of the *DSM-III-R* and *DSM-IV* personality disorders with the five-factor model of personality. In P. T. Costa, Jr. & T. A. Widiger (Eds.), *Personality disorders and the five-factor model of personality* (pp. 41–71). Washington, DC: American Psychological Association.
Wiggins, J. S. (1962). Strategic, method, and stylistic variance in the MMPI. *Psychological Bulletin, 59*, 224–242.
Wiggins, J. S. (1966). Substantive dimensions of self-report in the MMPI item pool. *Psychological Monographs, 80*, (22, Whole No. 630).
Wiggins, J. S. (1973a). Despair and optimism in Minneapolis (A review of J. N. Butcher (Ed.), Objective personality assessment: Changing perspectives). *Contemporary Psychology, 18*, 605–606.
Wiggins, J. S. (1973b). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley.
Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, 37*, 395–412.
Wiggins, J. S. (1984). Cattell's system from the perspective of mainstream personality theory. *Multivariate Behavioral Research, 19*, 176–190.
Wiggins, J. S. (1991). Agency and communion as conceptual coordinates for the understanding and measurement of interpersonal behavior. In W. Grove & D. Cicchetti (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl* (Vol. 2, pp. 89–113). Minneapolis, MN: University of Minnesota Press.
Wiggins, J. S. (1995). *Interpersonal Adjective Scales: Professional manual*. Odessa, FL: Psychological Assessment Resources.
Wiggins, J. S. (1996). Preface. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. vii–xi). New York: Guilford.
Wiggins, J. S. (in press). *Paradigms of personality assessment*. New York: Guilford Press.
Wiggins, J. S., Phillips, N., & Trapnell, P. (1989). Circular reasoning about interpersonal behavior: Evidence concerning some untested assumptions underlying diagnostic classification. *Journal of Personality and Social Psychology, 56*, 296–305.
Wiggins, J. S., & Pincus, A. L. (1989). Conceptions of personality disorders and dimensions of personality. *Psychological Assessment, I*, 303–316.
Wiggins, J. S., & Pincus, A. L. (1992). Personality: Structure and assessment. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 43, pp. 473–504). Palo Alto, CA: Annual Reviews.
Wiggins, J. S., & Pincus, A. L. (1994). Personality structure and the structure of personality disorders. In P. T. Costa, Jr. & T. A. Widiger (Eds.), *Personality disorders and the five-factor model of personality* (pp. 73–93). Washington, DC: American Psychological Association.
Wiggins, J. S., & Trapnell, P. D. (1996). A dyadic-interactional perspective on the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 88–162). New York: Guilford.
Wiggins, J. S., & Trapnell, P. D. (1997). Personality structure: The return of the Big Five. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 737–764). San Diego, CA: Academic Press.
Wiggins, J. S., & Trapnell, P. D., & Phillips, N. (1988). Psychometric and geometric characteristics of the revised Interpersonal Adjective Scales (IAS-R). *Multivariate Behavioral Research, 23*, 517–530.
Wiggins, J. S., & Trobst, K. K. (1997). Prospects for the assessment of normal and abnormal interpersonal behavior. *Journal of Personality Assessment, 68*, 110–126.
Winter, D. G. (1973). *The power motive*. New York: Free Press.
Woodworth, R. S. (1917). *Personal data sheet*. Chicago: Stoelting.

4.13 Observations of Parents, Teachers, and Children: Contributions to the Objective Multidimensional Assessment of Youth

DAVID LACHAR
University of Texas-Houston Medical School, Houston, TX, USA

4.13.1 ISSUES IN OBJECTIVE MULTIDIMENSIONAL MULTISOURCE ASSESSMENT 372
4.13.1.1 Overview 372
4.13.1.2 Multidimensional Measurement 372
4.13.1.3 Multisource Assessment 373
4.13.1.4 Psychometric and Technical Considerations 374
4.13.1.4.1 Scale stability and homogeneity 374
4.13.1.4.2 Scale construction methods and scale character 375
4.13.1.4.3 Evidence of scale and questionnaire validity 376
4.13.2 MULTIDIMENSIONAL MULTISOURCE RATING SCALES 377
4.13.2.1 Overview 377
4.13.2.2 Conners' Rating Scales-Revised 378
4.13.2.2.1 Conners' Parent Rating Scale-Revised 378
4.13.2.2.2 Conners' Teacher Rating Scale-Revised 379
4.13.2.2.3 Conners–Wells' Adolescent Self-Report Scale 379
4.13.2.2.4 Commentary 380
4.13.2.3 Child Behavior Checklist/4–18 and 1991 Profile (CBCL); Teacher's Report Form and 1991 Profile (TRF); Youth Self-report and 1991 Profile (YSR) 380
4.13.2.3.1 Child Behavior Checklist/4–18 (CBCL) and 1991 Profile 382
4.13.2.3.2 Teacher's Report Form (TRF) and 1991 Profile 383
4.13.2.3.3 Youth Self-report (YSR) and 1991 Profile 383
4.13.2.3.4 Commentary 384
4.13.2.4 Behavior Assessment System for Children 385
4.13.2.4.1 BASC Parent Rating Scales 386
4.13.2.4.2 BASC Teacher Rating Scales 386
4.13.2.4.3 BASC Self-Report of Personality (SRP) 386
4.13.2.4.4 BASC Monitor for ADHD 387
4.13.2.4.5 Commentary 387
4.13.2.5 Personality Inventory for Children, 2nd Edition (PIC-2); Personality Inventory for Youth (PIY); Student Behavior Survey (SBS) 388
4.13.2.5.1 Personality Inventory for Children, 2nd Edition (PIC-2); Personality Inventory for Youth (PIY) 389
4.13.2.5.2 Student Behavior Survey 397
4.13.2.5.3 Commentary 398

24

Assessment of Suicide Risk[1]

Ronald Stolberg
Bruce Bongar

Suicidal behavior is the most frequently encountered of all mental health emergencies (Schein, 1976). Several studies show changes in identity among members of high-risk groups (Boyd, 1983; Evans & Farberow, 1988; Maris, 1981, 1989; Roy, 1986; Shneidman, 1989; Weissman, Klerman, Markowitz, & Ouelette, 1989), and follow-up studies demonstrate that risk factors among attempters may be significantly different from those for the general population (Frederick, 1978; Maris, 1981, 1989). Also, studies suggest that rates of suicide among the mentally ill (Asnis et al., 1993; Hirschfeld & Davidson, 1988; Maris, 1981; Pokorny, 1964; Roy, 1986; Simon, 1988), and physically ill (Abram, Moore, & Westervelt 1971; Hirschfeld & Davidson, 1988; Maris, 1981; Meltzer, 1998; Reid, Mason, & Hogan, 1998; Rich & Runeson, 1992; Roy, 1986) far exceed that of the population as a whole. However, estimates of rate of death by suicide due to psychiatric disorders can vary greatly across such factors as country of origin (Roy, 1986), regional differences within a single country (Peterson, Bongar, & Netsoki, 1989), and the diagnostic criteria used for the sample. For example, estimates of the number of manic depressives who die by suicide range from 15% to 55% (Goldring & Fieve, 1984).

Clearly these variances mitigate against any single explanation of suicidal behavior or any simplistic approach to treatment (Bongar, Peterson, Harris, & Aissis, 1989). Indeed, the diverse explanations for suicide include biological predisposition (Arana & Hyman, 1989; Asberg, Bertillson, & Martensson, 1984; Brown et al., 1982; Mann & Stanley, 1988a; Mann, Oquendo, Underwood, & Arango, 1999; Roy, Nielsen, Rylander, Sarchiapone, & Segal, 1999; Stanley & Mann, 1988; van Praag, 1986); reaction to humiliation, helplessness, hopelessness, and guilt (Beck, 1989; Beck, Brown, Berchik, Stewart, & Steer, 1990; Shneidman, 1986b; Young et al., 1996); manipulation on the part of the patient (Gunderson, Kolb, & Austin, 1981); an escape from physical or psychological pain or an expression of violent rage (Nemiah, 1982); reaction to separation from family or from the loss of a loved one (Richman, 1986); and eroticization of death itself or an aesthetic completion of patriotic sacrifice (Lifton, Shuichi, & Reich, 1979).

A recent trend is to focus on new empirically based strategies in the assessment of suicide potential Bongar (2000). Some researchers have turned their attention away from making predictions of patient behavior to clinicians' own

views of critical factors when assessing for suicide (Bongar, 1991; Bruno, 1995; Greaney, 1995: Mahrer, 1993), attempting to identify the standards of care for this practice situation. This approach seeks to bridge the study of patient characteristics with that of clinician education, training, and experience in hope of describing reasonable and prudent practitioner behaviors. A profile of these behaviors not only is useful in identifying the standard of care, but can also help identify professional myths and deficiencies in practice, both essential to training and education efforts.

The assessment of suicide risk remains among the most important, complex, and difficult tasks performed by clinicians (Bongar, Maris, Berman, Litman, & Silverman, 1993; Chemtob, Bauer, Hamada, Pelowski, & Muraoka, 1989; Motto, 1991). Because suicidology literature is voluminous, diverse, and sometimes contradictory, clinicians may have difficulty determining the relative importance of various factors when assessing the individual. Some researchers have emphasized particular charts of variables. Gutheil and Appelbaum (1982), for example, stated that the best approach to gathering data "focuses on previous psychiatric history; recent behavior or behavioral change; significant alteration of circumstances (e.g., loss of job); bizarreness of ideation or action; threats to self or others, or related behavior such as the purchase of poison, rope, or a gun; history of substance abuse and the like" (p. 52). These authors pointed out that the central clinical and legal concerns involve negligence in evaluation and in involuntary interventions (i.e., hospitalization). Brent, Kupfer, Bromet, and Dew (1988) noted that an accurate diagnostic assessment is necessary with regard to primary and comorbid psychiatric disorder, alcohol and drug abuse, personality disorders, and attendant medical disorders. In addition, they pointed out that a proper assessment mobilizes the families and significant others in order to improve compliance with treatment and decrease the chance of a relapse.

Fremouw, de Perczel, and Ellis (1990), as well as Bassuk (1982), noted that the assessment of a patient's potential suicide risk necessitates the gathering and weighing of a variety of information and data—and the importance of this particular assessment has led these psychologists to construct an impressive decision model that integrates and formalizes the steps for a thorough and reasonable decision about the risk for suicide for a particular patient. Somewhat like Bassuk's 1982 checklist system, their decision model involves seven steps for the psychologist:

1. *The collection of demographic information* (e.g., age, sex, race, marital status, and living situation) to determine whether the patient is in a high-risk or low-risk group.
2. *The examination of clinical and historical indicators* as the more specific information that increases or decreases the patient's risk for suicide (e.g., unique historical, environmental, and psychological features that a patient presents in the intake interview, or during ongoing therapy).
   a. Questions about general historical-situational factors (which begin to lead the clinician to be concerned about a higher risk).
   b. Very specific clinical indicators and warning signs that often are the precursors of an imminent attempt (e.g., having a definite plan, strong self-destructive impulses).
   c. Psychological indicators such as recent losses, depression-anxiety, isolation-withdrawal, hopelessness, disorientation-disorganization, alcohol and drug use, change in clinical features, suicide plan, and final arrangements for his or her own death.
3. *An initial screening for risk*, that is, after examining historical-situational, demographic, and clinical indicators, the clinician must make a decision about whether the potential risk for suicide warrants any further assessment. If there appears to be no risk, assessment and treatment proceed in a routine fashion. However, if there are risk factors in the demographic or clinical-situational-historical matrix, the clinician should assess the current risk for suicide through two

formats, initially through interview, and if indicated, by self-report.

4. *Direct assessment of risk* using:
   a. *The clinical interview* (which includes the patient's reasons for feeling suicidal, as well as his or her reasons for living) and, where risk appears mild, moderate, or unknown,
   b. *Assessment by self-report* (e.g., the use of standardized assessment instruments such as the Beck Depression and Hopelessness scales to facilitate a thorough understanding).
5. *Determination of the level of risk and the implementation of a response* (e.g., none-to-low risk, mild, moderate, high).
6. *Determination of the imminence of risk* (e.g., assessment and documentation of rationale, consultation).
7. *Implementation of treatment strategies* (e.g., intensified outpatient care, voluntary hospitalization, involuntary hospitalization, etc.).

Although the above model is an important contribution as a systematic decision-making tool for the practicing clinician (and may very well represent the future general shape of standard models of assessment/intervention for suicide potential), Motto (1989), in a review of general problems in suicide risk assessment, pointed out:

> That assessment of suicide risk is, indeed, a problem is attested to by the fact that to date we have no established and generally accepted procedure to guide us in this task. Innumerable decisions regarding risk are made and implemented every day—the job gets done—but how it is done is determined primarily by the skills and philosophy of the individual clinician. (p. 245)

Motto (personal communication, July 1998) also noted that the accurate assessment of imminent risk does improve with clinical experience, and that supervision in working with suicidal patients should be a crucial element in all clinical training programs—"there is just no substitute for experience." Rosenberg (1997) pointed out that the suicide prevention literature primarily focuses on action-based interventions (e.g., assessment of risk factors and the use of direct interventions, such as hospitalization) for preventing suicide and that it lacks specific strategies to deal with the patient's emotional pain associated with suicidal ideation. In response to this need, Rosenberg (1997) designed a suicide assessment and intervention training model.

The clinical training approach suggested here differs from present training in several important ways. First, a systematic training model was developed and sequenced with respect to expertise and cognitive psychology principles (see Rosenberg, 1997). Few current training models utilize principles gleaned from this body of literature. Second, this model provides training in basic assessment skills, whereas most other programs focus on teaching basic interviewing or specific counseling skills (e.g., asking open-ended questions and learning reflection-of-feeling responses). Third, affective or feeling-based intervention strategies (intended to address underlying thoughts or feelings of suicide) are highlighted in this approach and supplement action-based strategies that are more commonly used.

Simon (1988) remarked that providing a reasonable standard of care in assessing and diagnosing suicidal patients preempts "the very problematic issue of prediction of suicide for which standards do not exist. Psychiatrists have not been held legally liable for inaccurate predictions of suicide per se" (pp. 86–87). When discussing what clinicians can do to reduce their legal liability when treating a suicidal patient, Simon (1992) identified "failure to diagnose" (assess the risk of suicide) (p. 274) as one of the three broad categories of negligence of which clinicians should be aware. Only when they "have failed to collect necessary data and logically assess it in making a prediction of suicide have lawsuits against psychiatrists prevailed" (Simon, 1988, pp. 86–87).

However, Lewinsohn, Garrison, Langhinrichsen, and Marsteller, 1989, pointed out:

> Unfortunately, the low base rate of suicide makes it very difficult for screening instruments to predict accurately the degree of suicidal risk for any specific individual. Specifically, the low base rate results in a very large number of false positive assessments relative to the number of true positives, even when the most effective screening instruments are used. (p. 4)

More recently, however, Bongar (1991) has constructed a comprehensive and integrated decision checklist that draws upon and recognizes the important contributions of Beck and his colleagues' (1979) empirical model (e.g. hopelessness, helplessness, negative cognitions), Fawcett and colleagues' (1987) empirical model (e.g., chronic *vs.* acute, the social matrix, communication of intent), Hirschfeld and Davidson's (1988) epidemiological model, Simon's checklist model, Shneidman's psychological model, and Yufit's Suicide Assessment Team/Suicide Assessment Battery/Suicide Screening Checklist (SAT/SAB/SSC) protocols (1988). This clinical and legal formulation of a standard of care for the assessment of elevated risk (Bongar, 1991) follows the decision checklist tradition and puts forward a large number of specific steps and detection-decision points. However, practitioners should be cautious: Such approaches often offer a possible set of known risk factors and are often not meant to be definitive or exhaustive. Mental health professionals will almost certainly tailor the assessment of elevated risk to what Motto (1989) has shown to be the uniqueness of every decision on suicide probability. Fawcett and his colleagues (1987), using data from a large-scale prospective study of patients with major affective disorders, found that 25 patients (of a total of 954) committed suicide; hopelessness, loss of pleasure or interest, and mood cycling during the index episode differentiated the suicide group (Fawcett et al., 1987). These researchers noted that, although suicide is a relatively frequent event in depressed patients, it still has a statistically low base rate and, therefore, may be statistically unpredictable on an individual basis using cross-sectional measures. Fawcett (1988) did offer a model that, while acknowledging the limitations of the current knowledge base, enumerates a variety of both short-term (acute predictors) and long-term suicide risk features (chronic predictors).

In recognition of the methodological and clinical difficulties in predicting suicide, the core of this chapter on assessment of suicide risk will focus on the collection of critical data and decision points in the clinical formulation, detection, and documentation of *imminent and elevated risk* in the usual and customary populations seen in professional psychological practice (Bongar, 1991, 1993, 1998; Bongar, Greaney, & Peruzzi, 1998; Bongar, Peruzzi, & Greaney in press; Bongar, Maris, Berman, & Litman, 1992; Bongar et al., 1989; Brent et al., 1988; Evans & Farberow, 1988; Fawcett, 1988; Fremouw et al., 1990; Hirschfeld & Davidson, 1988; Gutheil & Appelbaum, 1982; Jacobs, 1989a, 1989b; Kreitman, 1986; Lesse, 1989; Lewinsohn et al., 1989; Linehan, 1987, 1999; Linehan, Rizvi, Welch, & Page, 2000; Litman, 1989; Maltsberger, 1988; Mann & Stanley 1988a, 1988b; Maris, 1981, 1986, 1988, 1989; Motto, 1986, 1989; Murphy, 1986, 1988a, 1988b; Perr, 1985, 1988; Peterson & Bongar, 1989; Pokorny, 1983, 1993; Pope, 1989; Rachlin, 1984; Robins, 1985, 1986; Robins & Helzer, 1986; Roy, 1986; Shneidman, 1984, 1985, 1986a, 1987, 1989; Simon, 1987, 1988; Swartz, 1988; Swenson, 1986; VandeCreek & Knapp, 1983, 1989). In view of publications that focus specifically on the issues of youth suicide and suicide among the elderly (Berman & Jobes, 1991; Mcintosh, Santos, Hubbard, & Overholser, 1994), the present chapter will emphasize more general personality, clinical, and legal issues with regard to the formulation of a standard of care for adult clinical populations.

A clinical approach to the assessment of suicidal patients:

1. Recognizes the probabilistic nature of risk detection versus risk prediction.
2. Acknowledges the strengths and limitations of the traditional diagnostic categories (e.g., DSM), as well as the various theories of psychotherapy and psychopathology.
3. Understands epidemiologic and clinical risk factors within specific groups, taking into consideration the data on both acute and chronic predictors of risk.
4. Remembers when consultation, supervision, and referral are necessary.

5. Integrates a careful clinical history, mental status examination, ongoing clinical evaluations, consultations, information from significant others, and data from psychological assessment and suicide risk estimators/scales into a broad-spectrum information-gathering procedure for systematic assessment and management of detected risk.

Recently, seven factors relevant to suicide risk were identified by Joiner, Walker, Rudd, and Jobes (1999): previous suicidal behavior, type of current suicidal symptoms, precipitant stressors, symptomatic presentation, self-control and impulsivity, predispositions, and protective factors. When present, these seven factors have different implications for overall risk of suicide and are influenced by the patient's status as either a suicide ideator, single attempter, or multiple attempter.

The Modified Scale for Suicidal Ideation (Joiner, Rudd, & Rajab, 1997) identified two factors, "resolved plans and preparation" and "suicidal desire and ideation," as highly indicative of increased suicide risk. "Resolved plans and preparation" includes the following symptoms: "a sense of courage to make an attempt, availability of means to and opportunity for attempt, duration of suicidal ideation, and intensity of suicidal ideation" (Joiner & Walker, 1999, p. 448). The authors suggest that a patient who presents with these symptoms should be considered as at least a moderate risk for suicide. The second factor, "suicidal desire and ideation," includes the following symptoms: "reasons for living, wish to die, frequency of ideation, wish not to live, passive attempts, desire for attempt, expectancy of attempt, lack of deterrents to attempt, and talk of death and/or suicide" (Joiner et al., 1999, p. 448). In the absence of symptoms associated with the "resolved plans and preparation" factor and the other five factors listed above, and if the patient is not a multiple attempter, then the patient would not be considered at risk for suicide. The patient's past history of suicide attempts is used in conjunction with specific current suicidal symptoms, such as "resolved suicide plans and preparation," and is interpreted in relationship with the other five factors to produce a model for assessment of suicidal risk.

> For example, a patient who expresses a wish to die, who talks of suicide, and who reports frequent suicidal ideation can be worrisome to the clinician. In the absence of symptoms from the resolved plans and preparation factor, however (and in the absence of multiple attempt status or complicating factors from other domains), these symptoms do not warrant a high-risk designation. As an additional example, a patient who expresses little desire for death and relatively infrequent suicidal ideation, but who senses high competence and courage to attempt to suicide, who has means and opportunity, and who reports details of a suicide plan, is at high risk, regardless of other factors. (Joiner & Walker, 1999, p. 448)

However, before proceeding, it is essential to emphasize once again that hard and fast actuarial data on the long-term prediction of attempted or completed suicide—predictions that can be directly translated to the emergent clinical moment—do not presently exist. Currently, "there are no pathognomic predictors of suicide" (Simon, 1987, p. 259). Further, we will be following a tradition set by Monahan (1981) in his superb treatise on the clinical prediction of violent behavior, another clinical dilemma suffused with uncertainty and a lack of hard and fast actuarial data, namely, that this chapter represents "nothing more (or less) than the professional judgment of persons experienced at the task of prediction" (Monahan, 1981, p. 101) and that clinicians would do well to follow Meehl's advice that, "when actuarial data do not exist, we must use our heads" (Monahan, 1981, p. 108).

RECENT EMPIRICAL EVIDENCE ON RISK ASSESSMENT PRACTICES

In 1990, the Risk Assessment Committee of the American Association of Suicidology (AAS) reported the results of their survey on suicide risk assessment procedures among a random sample of practicing psychologists, psychiatrists, and clinical social workers (including members and nonmembers of the AAS). Although the results of this survey should be interpreted with caution due to the small sample size ($n = 414$) and relatively low response rate (38%), they do provide at least an initial glimpse into the assessment practices of practitioners (Jobes, Eyman, & Yufit, 1990). Also of interest was the amount of assessment activity carried out by the respondents, many of whom appeared to be psychologists—they reported evaluating, on the average, 1.9 suicidal adolescents and 2.8 suicidal adults per month.

Specifically, Jobes, Eyman, and Yufit (1990) found that:

1. Suicide assessment instruments appear to be used infrequently and are rated as having limited usefulness. For adolescents, the most frequently used were the Beck Hopelessness Scale (BHS) (28%) and the Beck Suicide Intent Scale (SIS) (23%). For adults the most frequently used was BHS (34%) and SIS (28%). These two scales were rated as "somewhat useful" for both adults and adolescents.
2. Although traditional psychological tests are used more frequently, they are not rated as very useful. The most commonly used tests were the MMPI (47%), Beck Depression Inventory (46%), the Rorschach (44%), and the Thematic Apperception Test (42%). Of these, only the BDI received a rating of "somewhat useful."
3. With a few exceptions, there are few major differences in the approaches used to assess acute versus chronic risk, and limited differences in the assessment of adults compared with the assessment of adolescents. Instead, most clinicians clearly rely primarily on the clinical interview to assess suicide (specifically, on certain valued questions and observations).

Linehan (Linehan et al., 2000) updated her 1981 model for analyzing risk factors associated with suicidal behaviors in an effort to determine whether risk factors among patients diagnosed with a personality disorder differ from those among individuals in the general population. Linehan noted that although many of the risk factors are the same, a few factors are actually not significant for individuals with borderline personality disorder. Linehan (Linehan et al., 2000) also explored why age, marital status, and gender did not appear to indicate a heightened risk for suicidal behavior.

> The absence of an effect for marital status may be due to low variance; individuals meeting criteria for this diagnosis are less likely to be married. The absence of an effect for age may have to do with the follow-up period in the studies reported. None followed elderly individuals meeting criteria for borderline personality disorder. The absence of gender difference in suicide parallels findings reported by Kreitman (1977) showing similar rates of suicide for males and females with histories of repeated parasuicidal acts; a criterion for borderlines. (p. 12)

Recent empirical studies of psychologists' reported practice behaviors confirm the primary reliance on the clinical interview as the major risk assessment tool in clinical practice (Card-Strong, 1998; Greaney, 1995; Peruzzi and Bongar, 1999). New empirically based strategies may well be needed to approach the assessment of suicidal potential. Some researchers have turned their attention away from making predictions of patient behavior to clinicians' own views of critical factors when assessing for suicide (e.g. Bongar, 1991; Bruno, 1995; Greaney, 1995; Mahrer, 1993), attempting to identify the standards of care for this practice situation. This approach seeks to bridge the study of patient characteristics with that of clinician education, training, and experience in hopes of describing "reasonable and prudent" practitioner behaviors. A profile of these behaviors not only is useful in identifying the standard of care, but can also help identify professional myths and deficiencies in practice, both essential to training and education efforts.

To shed some light on the preferences of active practitioners and their primary reliance on interview and observational data, this chapter will examine the historical and contemporary literature on the use of psychological tests, suicide scales, and risk estimators, as well as the epi-

demiological and clinical data on risk factors. This leads to the importance of formulation of clinical judgment via important clinical observations and key elements in the clinical interview, as integral components of a model of comprehensive clinical assessment and risk management.

Psychological Testing

The psychological assessment of suicidal risk is a process fraught with personal uneasiness and anxiety on the part of the mental health professional. The burden is awesome, and the responsibility is frightening (Neuringer, 1974a). Yet as Neuringer remarked:

> If all of the methodological problems associated with valid assessment of suicidal risk can be overcome, then the occurrence of false negatives will be severely diminished. The reduction of the number of false negatives (i.e., suicidal people who are erroneously diagnosed as nonsuicidal) is the chief aim of self-destructive risk assessment techniques. It is, in essence, the task of saving lives. (p. 15)

Motto (1991) stated that because each individual is a unique entity, the nature of assessment and measurement is difficult. He noted that a measure and observation that may determine suicide risk in one person may have different significance or no relevance at all for another person. He concluded that if clinicians are given the opportunity to establish a level of trust that ensures candor and openness, then the clinician is in the optimal position to assess risk in individual cases.

Kaplan (1994) reported that there is a high level of agreement between patients disclosing the same information about suicidal behaviors on a self-report measure compared with face-to-face communications with a clinician in a clinical interview. Kaplan found the one exception is that many patients tend to disclose more information regarding recent suicidal ideation on self-report forms. He did not find any discrepancies between self-report and verbal communication on suicidal information that was historical in nature.

However, it should be emphasized that the assessment of a patient's risk for suicide should never be based on a single score or scale. A complete evaluation of risk factors, such as the patient's psychiatric diagnosis, substance abuse, previous suicide attempts, a family history of suicide, and current level of functioning, should be considered in conjunction with psychological assessment results (Bongar, 1992a; Maris, Berman, Maltsberger, & Yufit, 1992).

The Rorschach Inkblot Technique

Historically, the Rorschach technique was the most commonly used method for estimating the risk of suicide, although it has been supplanted by more sophisticated psychometric instruments such as the MMPI and various suicide lethality scales (Neuringer, 1974b). Neuringer noted that:

> There certainly does not appear to have been any particular determinant, sign, or set of signs, or content which appears to be associated with suicide under all or even most conditions, i.e., there is no pathognomic sign on the Rorschach for suicide which is so strong that it transcends all the various states of suicidality or the differing conditions under which test data was elicited. (Neuringer, 1974b, p. 88)

However, the Rorschach may still be a potent tool for assessing the risk of suicide, if it is used correctly. Meyer (1989) argued that Exner's (1978) research provided an elegant and effective tool for the prediction for suicide potential. He cited a large number of specific factors from Exner's seminal contribution *The Rorschach: A Comprehensive System* (Exner, 1978) (for example, FV + VF + V + FD is greater than 2; occurrence of a color-shading response; Zd is greater than or equal to 3.5; 3r + (2)/R is less than 0.30, experience potential is greater than experience actual; CF + C is greater than FC; S is greater than 3; X + % is less than 0.70; pure H is less than 2; P is greater than 8, or less than 3; R is less than 17). Meyer (1989) noted that, in general:

> A number of authors concur with Exner that the number of responses is low in suicidals, especially when they are depressed, that less integrated color responses are more common, and that the number of popular responses are either very high or low (Swiercinsky, 1985). Also, there is some indication that transparency responses (such as light bulbs) or cross sectional responses (such as X rays) are found more commonly in potential suicides than in other persons; and the greater the number of vista and shading responses, the greater the depression-based suicide potential. (p. 319)

Meyer (1989) also pointed out the importance of content that suggests decay or geographic depression as indicators of suicidal ideation. He noted that "responses that suggest hanging or drowning or other direct means of suicide should obviously alert the examiner to further consideration" (p. 319). Here he is in agreement with Neuringer (1974b), who suggested that the clinician rely most on content from the Rorschach. Neuringer wrote:

> If suicidal content appears on the Rorschach, it should be taken very seriously since its manifestation could be an indication that self-destructive behavior is close to the surface, and that the suicidal person may be using the test protocol as a medium of communication to the examiner about his intentions and feelings. (1974b, p. 91)

Meyer (1993) examined the effect of the repose frequency (R) on the suicide constellation to determine the effects of the constellation's ability to discriminate among groups when reports deviate from the average length. He determined that within the suicide constellation, longer protocols are significantly more likely to have more than two vista or form dimension scores, a color-shading blend, more than three morbid scores, and more than three space responses. He concluded that more frequent responses are associated with higher scores on the S-Con. Frank (1994) warned that if a person manifests suicidal ideation or behavior and has a Rorschach with indicators of poor form quality (CF = C > FC and X + % < 70), this individual should be considered at risk for suicide.

Eyman and Eyman (1991) maintained that although single signs on the Rorschach have been used to predict suicide potential (e.g., transparency responses and color shading), the results of studies on single signs have been mixed. There is some clinical support for the use of these variables to assess suicide risk but, thus far, the research is insufficient to allow a clinician to feel confident in their use (Eyman & Eyman; 1991).

A children's suicidal constellation was developed by Exner and his colleagues (Exner & Weiner, 1982). The purpose of developing a different suicide constellation for children aged 8 to 18 was that many of the items in the original adult suicide constellation contained variables that were developmentally normal for children and adolescents. Silberg and Armstrong (1992) noted that some of the features of Exner's adult are typical of adolescents in general, such as C + CF > FC (color-dominated responses exceeding form-dominated color responses) and S > 3 (using white space more than three times).

Silberg and Armstrong (1992) sought to determine the usefulness of the Rorschach in uncovering suicidal potential in hospitalized adolescents. They compared severely depressed nonsuicidal patients with severely depressed suicidal patients to isolate the factors unique to suicidal behavior and unrelated to diagnostic variables. They had 138 subjects with ages ranging from 12 to 18 and carrying a wide variety of diagnoses. They concluded that the adult suicidal constellation did not significantly discriminate among the groups. However, they found that four of the six features of the index selected 64% of suicidal subjects. This indicated that suicidal adolescents are painfully introspective (with FV + VF + V > 0), exhibit painful arousal (color-shading blends > 1), are morbidly preoccupied (MOR > 0), are impulsive (CF + C > FC), and have misperceptions of people (M− and distorted reasoning skills, weighted special scores > 9) (Silberg & Armstrong, 1992).

According to Eyman and Eyman (1991), no studies were able to support the children's scale's ability to discriminate suicidal from nonsuicidal children. Furthermore, Exner and Weiner (1982) concluded that the children's suicide constellation has limited clinician util-

ity, since it contains many variables that are normal for children and adolescents.

Perhaps the Rorschach can be most useful in the assessment of suicide risk if it is combined with appropriate and sound clinical judgment (Eyman & Eyman, 1991). Eyman and Eyman (1991) cautioned that suicide is too complex a behavior to be adequately captured by a single sign. They suggested that the configurational or constellation approach also shows promise for the use of the Rorschach for the assessment of risk. They noted that "characterological features can be found among people who are seriously suicidal that differentiate them from mild-attempters and these characteristics can be assessed using the Rorschach and the TAT" (p. 48).

Neuringer (1974b) further cautioned that the clinician should not make an inference about suicidal intention from the Rorschach by itself; that case history material, and data from other psychometric instruments, can help to maximize the accuracy of the decision-making process. He concluded that "the clinician should strive to have all the data possible available to him. The presence of previous suicidal attempts and threats should compound the danger associated with the presence of suicidal Rorschach content" (p. 91). For additional information on use of the Rorschach with the suicidal patient, the reader is directed to Exner (1978, 1986), Eyman and Eyman (1991), Meyer (1989) and Neuringer (1974b).

Minnesota Multiphasic Personality Inventory (MMPI & MMPI-2)

Clopton (1974), in an extensive review of the MMPI and the assessment of suicide risk, pointed to the findings of Dahlstrom, Welsh, and Dalhstrom (1972) that the degree of the person's depression, that is, an elevated score on Scale 2 (D) of the MMPI, was a mood state frequently associated with a preoccupation with death and suicide. He continued:

> The implication of high scores on scale 2 depend on other features of the MMPI and upon the behavior of the person taking the test. For instance, it is their conclusion (Dahlstrom et al., 1972) that suicidal risk is greater when a person's MMPI results show a significant elevation on scale 2 but his behavior does not give any indication of depression, and he denies depressive thoughts and feelings, than when the depression indicated by a scale 2 elevation is clearly reflected in the person's behavior. (p. 118)

Clopton also noted that "the one standard MMPI scale found most frequently to differentiate suicidal and nonsuicidal groups is scale 2" (p. 129). Meyer (1989) agreed, and stated that the prototypical pattern for suicidal individuals is the 2-7/7-2 combination on the MMPI. He felt that this code specifically indicates the presence of suicidal ideation, and that whenever "2 is elevated above 80T, the clinician should pursue the question of possible suicidal ideation" (p. 317).

Meyer pointed out that the critical items on the MMPI concerning suicide should always be carefully checked and that the likelihood of suicidal ideation "being actualized increases as scores on scale 4, 8, and 9 rise, reflecting greater loss of control over impulses, a rise in energy available for behavior, and an increasing sense of isolation and resentment toward other people" (p. 317), respectively. Specifically, a rise in Scale 8 indicates that suicidal patients may actually kill themselves whether or not they want to succeed; their impaired judgment may bring about an inadvertent suicide. He also noted that people can become suicidal following a severe loss in their psychosocial world and that a high spike on Scale 2 is characteristic of these circumstances. Meyer believes that the "2-4/4-2 code is more likely to reflect a manipulative suicide. Where both scales 4 and 6 are elevated, in addition to at least a moderate elevation on scales 2 and 7, repressed anger and interpersonal hostility are as basic to a suicide attempt as manipulation" (p. 318).

Craig and Olsen (1990) compared drug addicts without a history of suicide attempts to drug addicts with a history of suicide attempts. They found that suicide attempters were characterized by higher levels of maladjustment (mostly in the areas of depression) and feelings of alienation, were more prone to utilize projection and externalization, and were more emotionally withdrawn. Significant differences were found on scales F, K, Depression, Hysteria, Psychopathic Deviance, Masculinity/Femininity, Paranoia, Psychasthenia, Schizophrenia, and Social Introversion.

Except for K, the suicidal attempters obtained higher mean scores. The modal profile of the suicide attempter group (excluding scales Mf and Si) had an average T-score elevation of 73.12 compared with 66.12 of the nonsuicidal group. In addition, suicidal ideators scored highest on the Pd scale compared with nonattempters and attempters. Suicide attempters scored higher on content scales DEP, ORG, and MOR compared with the no-history of attempt group. Attempters and ideators scored higher on the FEM content scale compared with the no-history group. Codetypes found to be unique to the suicide attempt group were 1-8/8-1, 2-9/9-2, 4-3, and 5-9. Codetypes that were specific to the no-history group were 2, 5, 9, 1-9, 6-2, and 8-9. In the suicide attempt group, 42% of the codes comprised the 2-4/4-2, 4-9/9-4, 4-8/8-4 and 6-8/8-6 codetypes.

The MMPI-2 is the most widely used instrument for assessing psychopathology in clinical practice (Greene, 2000). Inconsistent findings among retrospective comparisons of suicide attempters and nonattempting comparison groups have led some researchers (e.g., Eyman and Eyman, 1991) to conclude that, despite considerable research effort, no MMPI item, scale, or profile configuration has been found to consistently differentiate suicidal and nonsuicidal patients. They do concede that the restandardized MMPI-2 may provide more valid indicators of suicidality.

With the release of the MMPI-2, there is now available to practitioners the Koss-Butcher Critical Item Set Revised, listing 22 items that are related specifically to depressed suicidal ideation. However, Butcher (1989) noted that these critical items are not "designed to operate as scales. They are used to highlight item content that might be particularly significant in the individual's case. As sources of clinical hypotheses, the critical items might be used to key the clinician into problem areas or concerns the patient may have" (Butcher, 1989, p. 17). (For additional information on the specifics of these critical items, the reader is directed to the work of Butcher, Dalhstrom, Graham, Tellegen, & Kaemmer, 1989; and Butcher, 1989).

Although the majority of studies investigating the utility of the MMPI instruments during suicide assessment focused on the original version of the test, two recent studies evaluated the utility of the revised version of the test (the MMPI-2: Butcher et al., 1989) for the assessment of suicide potential. Sepaher, Bongar, and Greene (1999) found that approximately 20% of psychiatric patients endorsed items 506 and 520 among nine different well-defined MMPI-2 codetypes. These items were referred to as the "I mean business" items by Sepaher et al. because both directly inquire about a recent history of suicidal ideation or intent. Glassmire, Stolberg, Greene, and Bongar (2001) found that these and four additional similar items (150, 303, 524, and 530) were endorsed often by psychotherapy patients who had recently denied a history of suicidality during an intake interview. These authors found that these six items could be grouped together to provide a highly face-valid and internally consistent scale of self-reported suicide potential. The reader is referred to this article by Glassmire et al. for a table containing linear T-score conversions for this scale based on a large clinical sample.

Additionally, a recent study (Stolberg, Glassmire, Ricci, Greene, & Bongar, 1999) designed to establish practice guidelines and a standard of care for the use of the MMPI-2 with suicidal patients examined how a select group of specialists used the MMPI/MMPI-2 to assess risk. Delphi methodology was used, which resulted in expert consensus on the following questions: (1) How (if at all) do you use the Validity Scales to assess for suicide potential? (2) How (if at all) do you use the Clinical Scales to assess for suicide potential? (3) Which specific items, or questions, if any, do you use when assessing for suicide potential? (4) "How (if at all) do you use other Scales (e.g. Content, Supplementary, Research Scales) to assess for suicide potential?

Finally, Osborne (1985), after reviewing the

use of the MMPI in psychiatric practice, argued that the MMPI is best used within the context of other information-gathering techniques and should be viewed as providing the clinician with hypotheses that can be verified with other methods. Such other methods could include the use of other psychological tests, suicide scales and risk estimators, and a comprehensive clinical interview and history (Hendren, 1990).

16 PF

Meyer and Deitsch (1996) noted that on the 16 PF, suicide attempters appear consistently to show themselves as shy, tense, suspicious, expedient, emotionally unstable, apprehensive, self-sufficient, somewhat introverted, and extremely anxious: "Translated into 16 PF scales, this means high scores on Q4, O, 1, Q2 and lower scores on G, H, and C. Furthermore, repeat attempters typically score lower on scales Q3 and C than do first-time attempters, indicating less stability and more impulsivity" (p. 318).

Millon Clinical Multiaxial Inventory (MCMI and MCMI-II)

McCann and Suess (1988) examined the Millon Clinical Multiaxial Inventory (MCMI) profiles of 131 psychiatric inpatients to identify 1-2-3-8 codetypes on the basic personality scales. This profile reflects a schizoid, avoidant, dependent, and passive-aggressive blend of personality traits. Approximately 25% of the subjects obtained the 1-2-3-8 codetype. The authors' analyses of the diagnoses and clinical records of these patients suggested that this profile reflects an affect/mood disturbance with prominent depressive features and suicidal ideation suicide attempt prior to admission to the hospital. Summarizing, Meyer and Deitsch (1996) wrote of the MCMI that:

> As might be expected, elevations on the MCMI of 85 or higher on D and CC, and a high score on the Desirability Scale would be indicative of suicide potential (Choca, 1992). If N begins to increase and a very high D begins to decrease on repeated testing, the patient may be acquiring the necessary resolve and energy to complete the act; thus suicide potential may be high. Those personality patterns most likely to exhibit suicide potential are indicated by elevated scores on 2, 8A, 5, and/or C. These patterns are the most inconsistent and they are particularly susceptible to poorly developed support systems. High scores on scale 8A are particularly associated with manipulative suicide potential. (pp. 382–383)

McCann and Gergelis (1990) compared MCMI-II profiles of 40 psychiatric inpatients admitted for suicidal ideation with those of 40 patients admitted for a suicide attempt. They concluded that suicide attempters scored significantly higher on Scale Y (Desirability) compared with ideators. Elevations on Scale Y indicate suicide attempters have a higher tendency to deny problem areas, generate socially desirable response sets, and conceal psychological and interpersonal difficulties. Ellis, Rudd, Rajab, and Wehrly (1996) cluster-analyzed MCMI scores from 299 suicidal psychiatric outpatients. Four personality profiles emerged from their cluster analysis: negativistic/avoidant/schizoid, avoidant/dependent./negativistic, antisocial, and histrionic/narcissistic.

Thematic Apperception Test, Bender-Gestalt, and Other Selected Tests

After a review of the literature on the Thematic Apperception Test (TAT) as a diagnostic instrument for the assessment of suicide risk, McEvoy (1974) concluded that the literature on the use of the TAT as an estimator of suicide risk is clearly disappointing. He stated that "the literature is sparse and not easily compared for purposes of generalizations. Perhaps the only general conclusion is that the test has not proved to be useful for this purpose" (p. 102).

Eisenthal (1974) evaluated the application of five psychological tests to the task of assessing suicide risk: the Bender-Gestalt, the Rosenzweig Picture-Frustration Study, the Sentence Completion method, the Hildreth Feeling and Attitude Scale, and the Maslow Social Personality Inventory. He noted that the first three of these instruments are commonly used in clinical assessment. The last two instruments were selected for review because they were directed to two relevant aspects of suicidal behavior, feelings and self-concept. Eisenthal found two unsettling facts after his review. First, there is sparsity of work on this problem. Second, there is a lack of systematic evaluation-centered studies among the published research. He concluded that "the clinician shopping for tests with demonstrated utility in assessing suicidal risk will have to look elsewhere" (p. 146).

Beck Depression Inventory (BDI) and Beck Hopelessness Scale (BHS)

The revised Beck Depression Inventory (BDI) consists of a 21-item instrument designed to assess the severity of depression in adolescents and adults (Beck & Steer, 1987). As Beck and Steer (1987) themselves observed, during the last 26 years the "BDI has become one of the most widely accepted instruments in clinical psychology and psychiatry for assessing the intensity of depression in psychiatric patients (Piotrowski, Sherry, & Keller, 1985) and for detecting possible depression in normal populations (Beck, Steer, Kovacs, & Garrison, 1985)" (p. 1). A BDI total score can provide an estimate of the severity of the overall depression. Yet Beck and Steer stressed that it is also important to attend to specific item content:

> In particular, special attention should be given to those symptoms relevant to suicide ideation (p. 7). Beck, Steer, Kovacs, and Garrison (1985) have pointed out that that the BDI's Pessimism item (Item 2) "was nearly as predictive of eventual suicide in 211 suicide ideators as the 20 item Hopelessness Scale (HS)" (Beck, Weissman, Lester, & Trexler, 1974). Patients admitting to suicide ideation (Item 9) and hopelessness (Item 2) with ratings of 3 or 4 should be closely scrutinized for suicide potential.
>
> It also is important to observe the overall pattern of the depression symptoms that the patient is describing. The BDI reflects not only cognitive and affective symptoms, but also somatic and vegetative symptoms. For example, some suicidal patients will not express suicide ideation, but have actually stopped eating and sleeping. (Beck & Steer, 1987, pp. 7–8)

Because scores on the Beck Hopelessness Scale (BHS) produce only an estimate of the overall severity of hopelessness (i.e., the severity of an individual's negative attitudes about the future), Beck and Steer (1988) pointed out that it is clinically important to pay attention to other aspects of psychological functioning, in particular the patient's levels of suicidal ideation and depression. They noted the study by Beck and co-workers (1985) that reported that BHS scores of 9 or more were predictive of eventual suicide in 10 out of 11 depressed suicide ideators who were followed for 5 to 10 years after discharge from the hospital, pointing to earlier findings that hopelessness has repeatedly been found to be a better predictor of suicidal intention than depression per se (Beck, 1986). Beck and Steer (1988) continued: "Therefore, patients describing moderate to severe levels of hopelessness should be closely scrutinized for suicide potential" (p. 6).

Furthermore, Beck and colleagues (1985) found that, in a 10-year prospective follow-up of 165 psychiatric patients who had been hospitalized for suicidal ideation, hopelessness was highly predictive of eventual suicide. Of the 11 patients in their study who went on to complete suicide, 10 (90.9%) had BHS scores greater than or equal to 9. Beck and his colleagues recently extended this investigation to an outpatient population and found that in a prospective study of 1,958 outpatients seen at the Center for Cognitive Therapy, hopelessness (as measured on the BHS) was significantly related to completed suicide (Beck, Brown, Berchick, Stewart, & Steer, 1990). Beck and colleagues specifically found that

> A BHS cutoff score of 9 or above identified 16 (94.2%) of the 17 eventual suicides, thus replicating a previous study with hospitalized patients. The high risk group identified by this cutoff score was 11 times more likely to commit suicide than the rest of the outpatients. The BHS thus may be used as a sensitive indicator of suicidal potential. (p. 190)

Bongar (1991) reviewed research by Beck and colleagues and concluded that a score of 9 or more on the BHS is a useful cutoff for elevated risk. However, it is important to note that Beck and Steer (1988) also gave several case examples to demonstrate the complexities involved in using the BHS and BDI as predictors of suicide during therapy. They cited the case of a patient

who, at the time of his evaluation, showed on his self-report and interview report the signs of severe depression, but no suicidal ideation.

> His BDI score, for example, was 45. He endorsed item 9 as "I don't have thoughts of killing myself." Similarly, the Scale for Suicide Ideation did not indicate current suicidal thoughts or plans. However, his BHS score was 20. He came in for three consecutive sessions, cancelled the fourth, and did not show up for the next scheduled appointment. Scores on the BDI during the three sessions he had kept dropped from 45 at admission to 35, 37, and 37, respectively. During the sessions the patient denied verbally and on self-report forms any suicidal thoughts or plans. However, his BHS remained high (again a score of 20). (Beck & Steer, 1988, p. 22)

The case of this patient, who killed himself 3 days prior to the next scheduled appointment, demonstrates that "in the presence of a high BHS and dropping BDI, a psychotherapist should be alert to the possibility of a suicide attempt" (Beck & Steer, 1988, p. 22). Here, Beck and Steer noted that the BHS accurately reflected suicidal risk "whereas the patient's self report of suicidal ideation or suicidal item on the BDI did not" (p. 22). Beck and Steer also emphasized the need to inquire about reasons for any specific hopeless responses to BHS items, as the clinical exploration of these responses may lead to the patient's acknowledging suicidal wishes. They noted that the BHS is also a valuable tool for indicating that the patient is still pessimistic (and possibly suicidal) even though significant improvement in symptoms has occurred. By discussing the specifically endorsed items on the BHS with the patient, the therapist can pinpoint some of the particular situational and psychological factors contributing to a therapeutic impasse. Focusing on specific relevant items on the BHS can help to break up a pervasive hopelessness and foster collaboration (Beck & Steer, 1988, pp. 24–25).

For the purposes of risk detection, a score of 9 or higher on the BHS is of particular importance, because the BHS attempts to identify the "potential for fatal suicide attempts and not the behavior itself. . . . In interpreting the results of the present studies, hopelessness may best be construed as a risk factor, perhaps analogous to a history of smoking or elevated blood pressure as a predispositional factor in heart disease" (Beck et al., 1990, p. 194). Young and colleagues (Young et al., 1996) reported that stable levels of hopelessness over time may be more predictive of suicide attempts than a high level of current hopelessness at any one point in time, in patients with remitted depression.

Recently, Beck, Brown, Steer, Dahlsgaard, and Grisham (1999) designed a study to calculate odds ratios and to determine the usefulness of using the BHS for suicide ideation at its worst point in the patient's life (SSI-W), for measuring current suicide ideation (SSI–C), and for identifying patients at higher risk for suicide. They found that:

> patients who scored in the "higher risk" category for *suicide ideation at its worst point in their lives* had 14 times higher odds of committing suicide than patients who scored in the "lower risk" category. In contrast, patients who scored in the "higher risk" category for current suicide ideation of for hopelessness had approximately 6 time higher odds of committing suicide. (p. 7)

The SSI-W was found to be a significant predictor of eventual suicide and there was a "significant trend for the unique contribution" of hopelessness, as captured by the BHS. Current suicide ideation (SSI-C) was not found to be a significant predictor of suicide. This study indicated that clinicians should assess for the severity of past suicide ideation and for stable levels of hopelessness over a continued period of time as opposed to focusing solely on the patient's current suicidal ideation (Beck et al., 1999). It was noted that outpatients with high scores on the SSI-W and consistently high scores on the BHS who do not respond to therapy should be considered as a high-risk group for eventual suicide (Beck et al., 1999).

ASSESSMENT OF SUICIDAL IDEATION, INTENT, AND BEHAVIOR

A number of years ago, the Suicide and Suicidal Behavior Program of the Mood, Affective, and Personality Disorders Research Branch of the Division of Clinical Research at the National Institute of Mental Health convened a workshop to discuss and address "common problems and issues in research methodology with emphasis on developing a common core of assessment tools and standardized terminology (e.g., suicide versus attempted suicide)" (Alcohol, Drug Abuse, and Mental Health Administration, 1989, pp. 1–81) for use in the study of suicide. Also, the Child and Adolescent Disorders Research Branch of the Division of Clinical Research commissioned an exhaustive study that reviewed scales suitable for epidemiological and clinical research for the assessment of suicidal behavior in adolescents (Lewinsohn et al., 1989). Although these activities are certainly an important step toward promoting comparative nationwide clinical research efforts for the study of suicide via the use of a standardized set of assessment instruments, it is important to note that these recommendations and workshops have as their purpose the optimization of efforts toward collaborative and comparative clinical research, rather than the setting of *standards for frontline clinical practice.*

This ambitious attempt to standardize the assessment of suicide risk for the purposes of clinical research follows a long tradition of efforts at estimating and assessing the risk of attempted and completed suicide. These efforts have included assessment protocols that contain both standardized psychological tests and suicide risk estimator and lethality scales.

Motto (1989) noted that methodological and practical problems have plagued the development of scales of suicide risk

> to the point of discouraging even devoted and experienced workers in the field of suicide prevention. . . . These obstacles have been small samples, limited data, a low base rate, nongeneralizabilty of critical stressors, the individual uniqueness of suicidal persons, unknown and uncontrollable variables that contribute to outcome, ambiguity of outcome (e.g., "suicidal behavior"), and problems of demonstrating reliability and especially validity. (p. 249)

Lewinsohn and colleagues (1989), after comprehensively evaluating all available assessment instruments used to study suicidal behaviors in adolescents and young adults for the NIMH, concluded that:

> Perhaps the most fundamental critique we wish to level at the currently existing instruments is that insufficient attention has been paid to issues of validity. . . . The dearth of this type of information might also explain why researchers using different instruments to measure suicidal behavior have reported contradictory findings. . . . The success of the various instruments to determine which individuals are at high risk for suicidal behavior or other forms of self-destructive action has not been determined. (pp. 97–98)

As can be seen from the ongoing efforts of the NIMH, there continues to be enormous interest in the development of suicide risk scales and estimators. Motto (1989) stated that historically the only well-known measure was a little-used suicide risk scale of the MMPI and that contemporary efforts at scale construction began in 1963 when the Los Angeles Suicide Prevention Center developed a special scale for assessing callers to their center (Farberow, Helig, & Litman, 1968). Motto noted that this scale has been widely used by suicide prevention and crisis centers.

The Los Angeles Suicide Prevention Center Scale focuses on demographic and clinical characteristics of patients. Bassuk (1982) noted that, in this regard, it is similar to Tuckman and Youngman's 1968 Scale for Assessing Suicide Risk (SASR). The SASR is a scale used to identify among suicide attempters those persons with a high potential to commit suicide. However, citing problems of replication, Eyman and Eyman (1991) conclude that the "scale [SASR] is not useful for a psychiatric population and [we] also believe its use is premature in a general population as the original finding has not been replicated" (p. 23).

Also, the Index of Potential Suicide (Zung, 1974), which measures potential suicide risk and includes both clinical and social-demographic variables, has been shown to distinguish adequately between patients with no suicidal behavior and those with a history of suicidal behavior. However, this scale does not seem to have good predictive validity (Eyman & Eyman, 1991).

With regard to the evaluation of an actual suicide attempt scale, there have been several important developments. The Firestone Assessment for Self-Destructive Thoughts (FAST) was developed by Firestone and Seiden (1990) and is based on the concept that suicide and self-destructive behavior are influenced by an inner "voice" (e.g., a negative thought process). The voice process represents a well-integrated pattern of thoughts, attitudes, and beliefs, antithetical to self and hostile toward others, that is at the core of the patient's self-limitations and self-defeating actions. The voice varies along a continuum of intensity ranging from mild self-criticism to angry self-attacks and suicidal thoughts.

The FAST is a self-report questionnaire designed to be used as a screening instrument (Firestone & Firestone, 1996). The patient endorses frequency of negative thoughts on a 5-point Likert-type scale. These self-destructive thoughts can be conceptualized along the Continuum of Negative Thought Patterns and help clinicians direct their interventions toward the areas where their patients are experiencing psychological pain (Firestone & Firestone, 1996; Firestone & Seiden, 1990). Knowledge of where a patient score falls on the continuum can also assist clinicians in identifying patients who are at increased risk for suicide. Table 24.1 presents some of the items on the FAST.

The Suicidal Ideation Questionnaire (SIQ) (Reynolds, 1987, 1988), the Adult Suicidal Ideation Questionnaire (ASIQ) (Reynolds, 1991), and the Suicidal Behavior History Form (SBHF) (Reynolds & Mazza, 1992) were developed in response to the growing demand by clinicians for an effective and efficient way to assess for suicide ideation and systematically obtain information on a patient's history of suicidal behavior (Reynolds, 1987, 1988, 1991; Reynolds and Mazza, 1992). The ASIQ is a derivation of the 30-item SIQ that was initially developed for the assessment of suicidal ideation in adolescents. The ASIQ consists of 25 items that measure a specific suicidal thought or behavior. The patient rates each ASIQ item on a 7-point scale, ranging from general wishes of death to specific thoughts or plans, which assess the frequency of occurrence within the past month. High scores indicate frequent suicidal thoughts. The SBHF provides a format for the evaluation and documentation of previous suicidal behaviors. It can be used with adults and adolescents.

TABLE 24.1. Selected Items from the Firestone Assessment of Self-Destructive Thoughts

Level 1: Self-depreciating thoughts of everyday life
You just don't fit in.
You're different from other people.
Level 2: Thoughts rationalizing self-denial
You're too shy to make any new friends.
Look at all you have to do.
You're never going to get finished.
You can't take any time off.
Level 3: Cynical attitudes toward others
Can't you see he (she) is just taking advantage of you?
He (She) is such an exploitative person.
You can't trust men (women). Just depend on yourself.
Level 4: Thoughts influencing isolation
Just stay in the background.
Don't make waves.
Just look normal, don't let anybody know how you really feel.
Level 5: Self-contempt: vicious self-abusive thoughts
You can't do anything right!
You're disgusting!
You're a failure, a total failure.
Level 6: Thoughts that support the cycle of addiction
Have another drink (cookie).
You need to relax.
Go ahead and drink, you deserve it.
Level 7: Thoughts contributing to a sense of hopelessness
Your future is hopeless; things will never get better.
You're a burden to your family.
They'd be better off without you.
Level 8: Giving up on oneself
Nothing makes any difference anymore.
Why go on living?
The world is not a place that you can live in.
Level 9: Injunctions to inflict self-harm
Step off the curb, walk in front of that car.
Smash yourself, get yourself off this earth.
Level 10: Thoughts planning details of suicide
Pick a time, find a time when nobody will notice you're missing.
You're so miserable; you can't stand it anymore.
Just get out of it.
Get the pills to do it.
Level 11: Injunctions to carry out suicide plans
Why don't you end it all?
Go ahead! It'll be over in a minute.
Just do it. Kill yourself!

Source: Firestone & Firestone, 1996.

One historically important scale is the Risk-Rescue Scale developed by Weissman and Worden (1974), which defines a suicide attempt as an "event where the risk of death is extremely high and probability of rescue is extremely low" (Shneidman, 1985, p. 20). Smith, Conroy, and Ehler developed an 11-point scale, the Lethality of Suicide Attempt Rating Scale (LSARS), to measure the degree of lethality of a suicide attempt. Eyman and Eyman (1991), in their review of the assessment of suicide using psychological tests, found that the LSARS was a valuable tool in assessing the lethality of intent and method from a previous suicide attempt.

However, despite the development of a number of alternative instruments (e.g., Beck, Schuyler, & Herman, 1974; Cohen, Motto, & Seiden, 1966; Poeldinger, Gehring, & Blaser, 1973; Weissman & Worden, 1972; Yufit & Benzies, 1973; Zung, 1974), such instruments remain primarily useful "as research tools rather than aids for front-line clinicians" (Motto, 1989, p. 249).

As Motto (1989) noted, an alternative approach to the search for generalized indicators of risk has been to examine precisely defined populations and settings and to develop scales that would be "situation-specific" (p. 249), as, for example, the attempt by Litman (1975) to provide:

> An estimate of risk for nineteen different patient populations, such as suicide attempters seen in the hospital (moderate risk), depressed alcoholic middle-aged male callers to Suicide Prevention Center (high risk), and young female suicide attempters (low risk). Subsequent work continues to focus on "clinical models," defined in terms of personality characteristics or clinical picture; for example, "stable with forced change," "alienated," "nice person," "alcohol abuse," "drug abuse" (Motto, 1977, 1979), or "suicide attempter" (Pallis, Barraclough, Levey, Jenkins, & Sainsbury, 1982; Pallis, Gibbons, & Pierce, 1984) . . . as with earlier studies, these reports generated interest among researchers but did not have a demonstrable impact on clinical practice. (p. 249)

To the appraisal contained in the above list, Eyman and Eyman (1991) would add the work of Buglass and McCulloch (1970) and their Scale for Predicting Subsequent Suicidal Behavior, as well as Miskimins, DeCook, Wilson, and Maley's 1967 Suicide Potential Scale.

One probable explanation for the lack of impact of such scales, collectively or individually, is that in their development, "little attention was paid to providing clinicians with a simple brief procedure that could be quickly translated into a clear indication of suicide risk" (Motto, 1989, p. 250). However, there have been recent attempts to construct clinically useful screening instruments for utilization by the clinician. The following examples are meant to be representative of this approach to the assessment of suicide, rather than an exhaustive list of all available instruments. (For a model of such lists, the reader is directed to Eyman and Eyman [1991], and to Lewinsohn et al. [1989].)

Fremouw and colleagues (1990) noted the utility of the Reasons for Living Inventory (Linehan, 1985), a self-report measure with 48 items, including items that reveal the belief systems of a patient that "may serve as mediators of suicidal behaviors" (n.b., psychometric properties are still under investigation, and cutoff scores for the scale are currently unavailable). They pointed out that this self-report measure has been found to discriminate between suicidal and nonsuicidal individuals; specifically, Linehan's (1985) findings indicated "the absence of strong positive reasons to live are most indicative of suicidal behavior" (Fremouw et al., 1990, p. 56).

Eyman and Eyman (1991) noted that the Hillson Adolescent Profile (Inwald, Brobst, & Morrissey, 1987), which contains 310 true/false items, appears to be a promising tool for distinguishing suicidal and nonsuicidal adolescents in school and offender populations. They also noted that the Suicidal Ideation Questionnaire (SIQ) (Reynolds, 1988), which was developed to assess adolescents' thoughts about death and suicide, may be more appropriate and useful for school populations than for psychiatric ones.

The Suicide Probability Scale (SPS) is a short self-report measure designed to assist in the assessment of suicide risk in both adults and adolescents. This 36-item self-report measure asks patients to rate their present and past behaviors on a 4-point Likert scale. Cull and Gill (1982) stated that:

> Assessment of suicide risk is reflected in three summary scores: a total weighted score, a normalized T-score, and a Suicide Probability Score which can be adjusted to accommodate different a priori base rates for particular clinical populations. . . . The SPS is intended solely as a screening instrument. It should not be used in isolation. Instead other methods such as clinical interview by trained psychiatric professionals should be used to supplement, corroborate, and investigate test results. (pp. 2–4)

This instrument, although serving as a useful model for interesting future directions in the development of assessment instruments, has a number of limitations. The intent of the scale is not particularly disguised, and it assesses an individual's reported feelings and behaviors only at one point in time. However, as the authors themselves recognized, further research on this instrument is needed to establish predictive validity, replicate findings with a wider range of representative samples, and assess the incremental validity of the SPS "in predicting suicidal behaviors beyond what could be predicted on the basis of commonly available patient demographic and clinical characteristics alone" (Cull & Gill, 1982, p. 61).

Beck and colleagues developed two important scales for the measurement of suicidal ideation and intent. These are the Scale for Suicide Ideation (SSI) and the Suicidal Intent Scale (SIS) (Beck, Kovacs, & Weissman, 1979; Beck et al., 1974). The SSI has been used in several studies to measure intensity, duration, and specificity of psychiatric patients' wishes to commit suicide (Beck, Steer, & Ranieri, 1988). The SSI is a 19-item instrument that a trained clinician may use to rate the severity of a patient's suicidal thoughts and plans on a 3-point scale that ranges from 0 to 2. Rather than employing cutoff scores, a clinician who detects any positive ideation on item 4 (Active Wish-to-Die) or item 5 (Passive Wish-to-Die) rates the patients on the remaining 14 items of the scale. The Center for Cognitive Therapy recommends that SSI total scores be used only as measures of suicide ideation in "true" ideators and not in general clinical populations. From a clinical standpoint, the Center for Cognitive Therapy considers an SSI score of 10 or higher as indicating suicidal risk and suggests that patients be followed closely.

Lewinsohn et al. (1989) recognized the SSI as a carefully developed, reliable instrument for measuring suicidal ideation. Beck and colleagues (1988) also reported that the SSI, in a self-report version, can be a reliable method for measuring the severity of suicide ideation in both outpatients and inpatients. However, they noted that "although the present study indicates that the SSI represents a valid and reliable method for rapidly estimating suicide ideation in psychiatric patients, it is not recommended that the self-report version of the SSI replace clinical interviewing as a method for evaluating a patient's suicide plans and thoughts. The self-report version represents another tool for multimethod assessment of suicide ideation" (p. 504).

Lewinsohn and coworkers (1989) concluded: "given that intentionality is an important construct in the study of suicide, the SIS represents the only rationally derived scale to evaluate suicidal intent" (p. 15). They cited Steer and Beck's (1988) belief that "the information elicited by this test can help clinicians judge how serious the attempt was and might be again, while noting that "its use is restricted to people who have made a previous attempt which is a high risk group" (Lewinsohn et al., 1989, p. 15). The SSI and the SIS represent important directions for future clinical research efforts and further the cause of combining clinical interview with clinical assessment protocols such that suicidal ideation and intent may be assessed more systematically. From a general clinical assessment perspective, Bassuk (1982) concluded that "by using the scale for suicidal ideation, the hopelessness scale, and the suicide intent scale, the interviewer can assess either directly or indirectly the seriousness of the patient's intent or subjective wish to die" (p. 29).

Hirschfeld and Russell (1997) designed an Algorithm for Assessing the Risk of Suicide to be used in conjunction with identified Risk Factors for Suicide and Attempted Suicide.

If the patient is found to be at risk for suicide, they recommend that the Algorithm for the Management of Suicidal Thoughts or Behavior be used by the clinician. The authors note that once treatment is started for patients identified as high risk there should be close follow-up and continued assessment, as "Some patients are at even greater risk for suicide for the first few days after treatment has been started" (Hirschfeld & Russell, 1997, pp. 912–913). For individuals with chronic suicidal ideation, the goal of the clinician is to change risk factors that can be modified in an effort to increase patient safety. The authors note:

> A person who is determined to kill himself or herself will probably prevail despite the best efforts of family members and health care professionals. However, the overwhelming majority of people who desire to kill themselves at one time will feel very different after improvement in their depression or after receiving help with other problems. (Hirschfeld & Russell, 1997, p. 914)

Motto, Heilbron, and Juster (1985) developed an empirical suicide risk scale for adults hospitalized due to a depressive or suicidal state. Their study of 2,753 suicidal patients prospectively examined 101 psychosocial variables. After a 2-year follow-up, 136 (4.94%) of the subjects had committed suicide. The authors employed rigorous statistical analysis, including a validation procedure, to identify 15 variables as significant predictors of suicidal outcome. Their findings were translated into a paper-and-pencil scale that gives an estimated risk of suicide within 2 years. Motto (1989) noted that instruments such as these can provide a valuable supplement to clinical judgment, as well as the kind of quantitative expression of suicide risk that represents to many clinicians "fine tuning" of their clinical judgment.

However, Clark, Young, Scheftner, Fawcett, and Fogg (1987) undertook a field test of Motto and colleagues' Risk Estimator for Suicide (Motto et al., 1985) that "raised questions" about the instrument though without invalidating the scale (Clark et al., 1987, p. 926). Clark and colleagues selected a subset of psychiatric patients with major or chronic affective disorders that corresponded to those in Motto's sample. The subjects in their sample exhibited distinctly lower suicide rates over a 2 year follow-up (2.4%) than the sample reported by Motto (4.9%). What the 1987 study does highlight is the critical need to understand the limitations of all such scales:

> . . . particularly the likelihood that suicide scales derived by multivariate analysis of a large number of clinical, psychosocial, and demographic variables may tend to be arbitrary and sample specific. Our impression is that empirically derived scales based on a single cross-sectional assessment are always difficult to validate. Repeated assessments over time on a broad array of clinical features may be necessary to develop an adequate and replicable prediction system. We propose that serial assessments which pay attention to the vicissitudes of clinical symptoms, changing life stress, and long-standing character structure in concert (K. Smith, 1985) would provide a better method of estimating suicide risk. (Clark et al., 1987, p. 926)

ASSESSING SUICIDE THROUGH STRUCTURED INTERVIEWS AND A PSYCHOLOGICAL BATTERY

Yufit (1988) proposed that the assessment of suicidal behavior is best conducted through the use of a Suicide Assessment Team (SAT). Such a team would be composed of a multidisciplinary staff of psychologists, social workers, nurses, and psychology graduate students, specially trained in the use of a focused screening interview format and other assessment techniques for the identification and evaluation of suicide potential. SAT assessment would involve three levels: a focused interview (Level I), this interview plus specialized rating scales (Level II), and an extended psychological assessment (Level III) including the above interviews and ratings as well as special psychological assessment techniques, termed the Suicide Assessment Battery (SAB).

The SAB would be used to make an extended evaluation of suicide potential beyond

interview ratings (Level I) and the scores on the specialized suicide rating scales (Level II). Before proceeding to Level III, the clinician already would have conducted the structured interview and used rating scales such as the Beck Depression Inventory, as well as the Risk-Rescue Scale and the Los Angeles SPC Assessment of Suicide Potential to evaluate the patient. As Yufit noted, Level III techniques "would most likely be used with inpatients who make suicidal threats or attempts, or where suicide can be inferred" (1988, p. 26).

One of the most important elements in the SAB is the Time Questionnaire (TQ). The TQ, a semiprojective personality assessment technique, has been found to correlate with suicide potential (Yufit & Benzies, 1979). The TQ has been administered to "over 1500 persons, including clinical and non-clinical samples, as well as matched sample populations; it has consistently differentiated high lethal suicidal persons from lower levels of suicide lethality and nonsuicidal persons as well as a variety of psychiatric diagnostic groups, on the basis of uniquely different time perspectives . . . the TQ is a key technique in the SAB" (Yufit, 1988, p. 27). Yufit (personal communication, April, 1990) added the Coping Abilities Questionnaire, a 15-item instrument that measures the range of coping ability (from excellent to minimal).

In addition to the Time Questionnaire and Coping Abilities Questionnaire, other possible key elements in a Suicide Assessment Battery include the following:

1. Q-sort set, a 22-item set representative of "descriptive items relating to variables often associated with suicidal behavior" (Yufit, 1988, p. 28)
2. Suicide Assessment Checklist, a 36-item checklist to provide a supplement to clinical judgment
3. Experience Inventory, in which the patient is asked to list 10 most important experiences (Cottle, 1976)
4. Motto Suicide Risk Assessment, an empirically derived instrument using significant items from a large-scale prospective study of a clinic instrument used to assess suicide (Motto et al., 1985)
5. Sentence Completion, a 32-item form with sentence stems especially selected to elicit affect related to morbid thoughts, self-destruction, hope, trust, the future, and so on
6. Draw-A-Person, DAP in the Rain, a variation of the "Draw-A-Person in the Rain" projective technique in which rain is an ambiguous stimulus . . . Scoring is subjective in nature, but the work of Machover and other exponents of DAP can be utilized" (Yufit, 1988, p. 30)
7. Autobiography, in which the patient is asked to write his or her life history in any way that he or she wishes
8. Rorschach, primarily used for associational content
9. TAT, quantitative analysis of story themes related to isolation, hopelessness, mistrust, morbid content, and future orientation
10. Object Relations Technique, a variant of the TAT that usually gives more elaborate data than the more structured TAT
11. Erikson Questionnaire, a multiple-choice instrument giving scores on Erikson's developmental model of stages related to intimacy-isolation, trust-mistrust, autonomy-shame, and doubt
12. Humor Test, a 104-item, objectively scored questionnaire that gives polar opposite scores on 13 factor analytically determined scales
13. Hope Scale, a Fawcett-Sussman scale of items relating to the evaluation of hope as personality variable

These 13 assessment techniques represent a broad array of available and useful measures from which a Suicide Assessment Battery can be derived, usually including five or six of the techniques (e.g., the core SAB battery might include the Suicide Checklist, Coping Questionnaire, Time Questionnaire, and Q-sort) (Yufit, personal communication, August, 1990).

Yufit (1988) pointed out that the use of the SAB, which includes a number of scales still in experimental form, "may be questioned, but they are considered very useful to trying to fill the lacunae and tap the nuances in the complex task of identification and assessment of suicide potential" (p. 32). He concluded that:

> Even in a psychiatric hospital setting, where psychiatric sophistication may be considered deep, there is a need for more *comprehensive* evaluation procedures of the complex behavior of suicide. At this stage of development, these techniques are not necessarily conclusive, nor are they often objective, but they very often do serve as important *guidelines* to assist in the identification and the assessment of the components of suicide potential. They should supplement clinical judgment not substitute for it. (p. 33)

Recently the Question, Persuade, Refer, and/or Treat (QPRT) was developed by Quinnett and Bratcher (1996) to help clinicians evaluate, counsel, and treat suicidal patients. This structured risk assessment interview can be used in conjunction with other scales, such as the BDI or the BHS. The QPRT is a "tactical interview guide" created to help the clinician obtain dynamic and critical information regarding the urgency, nature, and context of a suicidal crisis, as well as to assess the current level of immediate risk. A decision tree format is used to guide the clinician and patient in creating a crisis management plan.

In an effort to formalize and distill the elements of risk detection (as well as to try to increase the accuracy of diagnosing suicide risk), Yufit (1989) developed and field tested a new integrated instrument, the Suicide Screening Checklist (SSC). The SSC is used by a clinician or interviewer to assess an individual for the "purpose of identification of suicidal potential. Suicidal potential (or suicide risk) refers to the likelihood that such a person will engage in behavior that will directly or indirectly lead to self-destruction" (p. 4-129). The SSC is a screening instrument constructed from empirical data and utilizing known and presumed correlates of suicidal behavior, is designed to complement and improve "the validity of clinical judgment" (p. 4-139). In short, instruments such as the SSC (as well as batteries such as the SAB) may allow clinicians to supplement their own clinical intuition with a systematized approach to collecting assessment information (Yufit, personal communication, August, 1990). An accurate (and widely accepted) model for the assessment of elevated risk in the suicidal patient may require a future research effort that would involve a large-scale collaborative multicenter study designed to evaluate *all* of the existing assessment procedures for both efficacy and significance (Yufit, personal communication, August, 1990).

LIMITATIONS OF THEORETICAL ORIENTATION AND DSM-IV DIAGNOSIS

When clinicians select assessment criteria and then implement an appropriate intervention strategy with suicidal patients, they may find that traditional theories of psychotherapy, and traditional psychiatric diagnostic categories, are of limited practical value in precisely assessing suicidal risk. As Beutler (1989) pointed out, "the descriptive dimensions embodied in the current diagnostic system bear little relationship to the selection of the mode or frequency of psychosocial interventions . . . while it would be unthinkable in any practice of medicine for the mode of treatment to be independent of patient diagnosis, this is precisely the case in the assignment of psychotherapy modes and formats" (p. 272).

An integrated perspective on assessment and treatment of the suicidal patient must be maintained (Simon, 1988). For example, psychoanalytic approaches may tend to deemphasize discussion of the suicidal patient's condition with family members, or deemphasize an evaluation for the efficacy of organic therapies—in which case vital information can be lost. Conversely, biologically based approaches may not place enough emphasis on the need for an ongoing treatment relationship, that is, may diminish the role of the therapeutic alliance as an essential element of sound psychopharmacologic intervention. Finally, Simon (1988) noted the danger of medications being prescribed in lieu of the patient being seen more frequently.

Jacobs (1989b) asserted that the clinician must go beyond formal psychiatric diagnosis, because many suicides occur in individuals who

have not been labeled as psychiatrically ill. He cited Mack and Hickler's research (1981), which found: the problem of suicide cuts across all diagnoses; many of those who take their own lives are mentally ill, but some are not; some are psychotic, but most are not; and some act impulsively, but most do not. Jacobs (1989b) also noted that the next concern for the clinician with regard to assessment relates to an understanding of demographic and clinical risk factors.

Linehan and colleagues (2000) noted that the limited research associating Axis-II disorders with suicide is largely attributable to the fact that Axis-II disorders were not formally diagnosed before 1980. In hierarchical and principal diagnosis systems, Axis-II diagnoses are typically assigned only when an Axis-I disorder is absent. In this context, suicidal ideation or behavior is viewed as a symptomatic response to an Axis-I disorder. The high comorbidity of Axis-I and Axis-II disorders guarantees a significant underestimate of the prevalence of personality disorders.

Furthermore, personality disorders are typically represented by dimensional rather than categorical measures. That is, a specific number of criterion behavior patterns need to be present, and typically the greater the number of behaviors the greater the disorder. Linehan (2000) noted that:

> While a few studies attempt to examine the relationship of personality disorders and suicidal behavior using a more dimensional approach, the methods used have been haphazard. In the absence of consistent methodology, we are left with information obtained from categorical diagnostic systems. It should be clear, however, that this is not without its pitfalls. As Widiger and Frances (1987) point out, criterion behaviors are at times used as indicators of some underlying pathology (diagnostically), and at other times, they are used as operational criteria of a disorder (definitionally). Research, in many cases, does not sufficiently distinguish which criteria are definitional and which (if any) diagnostic. Problems with validity and diagnosis are thus, for the categorical classification system, inextricably intertwined. (p. 7)

Clark (1990) believed that there are a number of different types of suicide, based on different demographics, motives, or diagnoses. He thought that developing specific risk profiles for each of these different types of patients would be prudent. Clark and Fawcett (1992) advocated the formulation of diagnosis-specific suicide risk profiles. Clark (1990) provided four reasons for why diagnosis-specific profiles may be useful. First, he noted the relationship between mental illness and suicide (e.g., psychological autopsy studies). These studies indicate that most people who die from suicide suffered from some form of mental illness.

Second, Clark described follow-up mortality studies of patients with major affective disorders, alcoholism, and schizophrenia that indicated an elevated risk of death by suicide for each of these diagnoses. Third, he pointed out that different diagnoses (e.g., major depression, alcoholism, and schizophrenia) have different clinical presentations, courses, prognoses, and treatments, so the risks for suicide in patients with these diagnoses are also likely to differ. Fourth, Clark believed that the advent of structured interviews and well-delineated diagnostic categories would make reliable research of these categories possible.

Along these lines, Peruzzi and Bongar (1999) identified eight critical risk factors for suicide in patients with major depression: the medical seriousness of previous attempts, history of suicide attempts, acute suicidal ideation, severe hopelessness, attraction to death, family history of suicide, acute overuse of alcohol, and the experience of recent loss or separation. Additionally, Canapary (1999) found a two-tier series of suicide risk factors with an alcoholic population. The first tier of risk factors includes the seriousness of previous attempts, past suicide attempts, and the communication of suicidal ideation; the second tier includes feelings of hopelessness, a family history of suicide, and the current use of alcohol.

Suicide is not a psychopathological entity that is recognized by the *Diagnostic and Statistical Manual of Mental Disorders* (Lewinsohn et al., 1989). Indeed, major depression and borderline personality disorder are the only diagnoses in the DSM-IV for which suicidality is a criterion. Zubin (1974) noted that suicide is the final stage of a very complex process that encompasses a heterogeneous set of phenomena and that occurs in people who are anything but homogeneous (Lewinsohn et al., 1989, p. 1).

It is clear that the various therapies and procedures from the diverse schools of psychotherapy are applied across diagnostic groupings of patients. Beutler (1989) pointed out that the "rapid growth of integrative approaches to mental health treatment in recent years attests to the readiness of therapists to accept comprehensive theories of intervention which cut across disciplines, theoretical systems, and diagnostic labels to provide a comprehensive system for selectively assigning different treatment protocols" (p. 273). He cited the work of Cole and Magnussen (1966) and the more recent work on "operational diagnosis" by Cummings and VandenBos (1979). He also called for basing diagnostic and assessment decisions on a more disposition-focused assessment, that is, for the decisions about patient care to rest on a foundation of empirically derived relationships between patient characteristics, available environmental systems, and therapeutic effects. Such a focus on disposition, rather than merely being on symptoms, would keep the orientation on the "primary mission of enhancing the efficacy of care" (Beutler, 1989, p. 273). Beutler concluded:

> This method of selectively applying treatments to patients "represents a compromise between the intense matching of individual patients and therapists and the more global effort to match type of therapy to patient diagnosis." While still in its infancy, this approach to the problem of treatment selection is represented in the so-called "technical eclectic psychotherapy movement" (Lazarus, 1981; Norcross, 1986). Dispositional assessment proposes that diagnosis is a dynamic process that delineates the probability that an identified problem will respond to a given treatment or setting. An assessment procedure such as this will entail a sequential multistepped process of defining treatment propensities, needs, settings, enabling factors, and social contexts through which effective treatment occurs. (pp. 275–276)

In summary, at the level of a diagnostic and initial assessment of the suicidal patient, we believe that the psychologist should concentrate primarily on dispositional assessment and on the formulation and documentation of *risk detection* in the assessment and diagnostic stage, rather than on specific behavior prediction (Maris et al., 1992; Murphy, 1988b). However, the mental status examination, clinical interview, and DSM-IV diagnostics are essential and critical elements in a comprehensive clinical assessment process (see also Robins & Helzer, 1986; Black & Winokur, 1990) and, for the purposes of the present chapter, are important tools in the detection of suicide risk.

GENERAL FORMULATION OF CLINICAL JUDGMENT IN THE CLINICAL INTERVIEW

Maltsberger (1988) believed that there are five specific components in the general formulation of suicide risk:

1. assessing the patient's past responses to stress, especially losses
2. assessing the patient's vulnerability to three life-threatening affects—aloneness, self-contempt, and murderous rage
3. determining the nature and availability of exterior sustaining resources
4. assessing the emergence and emotional importance of death fantasies
5. assessing the patient's capacity for reality testing (p. 48)

He added that "what is sometimes called the 'formulation of suicide risk' offers the clinician a disciplined method for assessing suicide danger that integrates and balances the presenting clinical material from the patient's past history, his present illness, and the present mental status examination" (p. 48).

Another approach to the formulation of clinical judgment and the assessment of risk is a clinical checklist method of conducting the assessment (Bassuk, 1982). In this approach, one uses combinations of structured interviews, checklists, standard psychological instruments, and suicide risk scales and estimators to ensure a comprehensive evaluation. (For additional information on comprehensive assessment and interviewing strategies, see Hendren, 1990).

Yufit (1988) noted that the assessment of sui-

cide potential will often be accomplished through the initial focused interview (based on either a referral from the emergency room, inpatient service, or outpatient clinician). The Level I (focused interview) would explore:

1. The patient's conscious intent of actually ending his/her own life
2. Risk of rescue or possible interruption during suicide attempt
3. Degree of planning
4. Behavior level (e.g., threats, ideation, gestures, overt attempt)
5. Lethality of attempt made
6. Extent of physical injury and/or toxicity
7. Precipitant factors
8. Intensity of current life stress
9. History of previous attempts, gestures, threats, and ideation
10. Degree of depression
11. The patient's ability to relate to the examiner during the interview, overt dress and grooming, posture, degree of agitation, ability to discuss the problem, and so on
12. Changes in the patient's behavior during the initial contact
13. Overall psychological status (e.g., the examination of any DSM-IV disorders)

The majority of patients at the Center for Cognitive Therapy at the University of Pennsylvania carry a diagnosis of depression, and therefore hopelessness and suicidal ideation will often be ongoing therapeutic issues. The center's policy manual indicates that when the patient's suicidal ideation is either continuous or intermittent, that is to say consists of low to moderate levels of purely ideational symptoms, it can be dealt with as with any other depressive symptom; that is, primarily by the treating therapist with regular supervision (R. J. Berchick, personal communication, September 20, 1989). However, the center's special procedure for suicidal patients is:

> The therapist should maintain especially close monitoring of *all* suicidal thoughts as well as concomitant levels of hopelessness and depression. Regular administration of the first five items of the SSI (and the rest of the scale if any of these are positive), the BHS, and BDI at each session is strongly recommended, because these scales will provide a reliable means of monitoring any changes in these variables over time.

Richman and Eyman (1991), citing the previous work of K. Smith (1985), Eyman (1987), and K. Smith and Eyman (1988), have developed a model for understanding why a person chooses to commit suicide based on the data from the Suicide Research Project at Menninger's. This model posits three conditions for suicide:

1. a narrowly defined, unrealistic, and fragile identity
2. an event that jeopardizes the individual's identity
3. deficits in the management of affect and difficulties in problem solving (Richman & Eyman, 1991, p. 2)

Motto (1989) held that the central clinical task is to "determine and monitor the patient's threshold for pain (either physical or psychological)." This would take into consideration the person's pathology, strengths, and available defensive patterns. If the pain level exceeds the pain threshold (even briefly), Motto saw suicide as imminent. Therefore, the psychologist must carefully assess and monitor the patient's threshold for pain and estimate how close the current level of pain comes to it. "The better we know the patient, the more sensitive we can be to the influences that can alter these two critical determinants of a suicidal act. Treatment aims ideally at both raising the threshold by maturational development, and decreasing the pain level by providing emotional support and by resolution of pain-generating conflicts" (Motto, 1989, p. 254).

Shneidman (1987) likewise believed that:

> The central feature of suicide is pain, and that the key to suicide prevention lies in the reduction of that individual's psychological pain. All else—demographic variables, family history, previous suicidal history—is peripheral except as those factors bear on the presently felt pain. Ultimately, suicide occurs when there is the co-existence of intolerable pain, intense negative press, and extreme perturbation with perceptual constriction and an irresistible penchant for life-ending action. (pp. 176–177)

Clinicians must also, according to Shneidman (1984, 1986a), exercise extreme caution with any patient who is perturbed and who has a lethal means available. This would include clinical work with patients with poor impulse control who are in crisis and are unable to decrease their level of perturbation in the therapeutic encounter. Shneidman (1987) presented a theoretical cubic model of suicide that includes the combined effects of psychological pain, perturbation, lethality, and what he termed "press" to attempt to identify those individuals most at risk for suicide. Here "press" is similar to what Murray (1938) called pressure—that is, those aspects of the inner or outer world or environment that touch, move, impinge on, or affect an individual, and to which he or she reacts. "Press" can be either positive or negative.

Practically speaking, Pokorny (1983) noted that the identification and care of the suicidal patient in clinical practice is made up of a *sequence of small decisions*, a point we wish to underscore. Murphy (1988a) expanded on this dictum by noting that:

> The first decision might be based on some alerting sign or clinical configuration, and the decision would be to investigate further. After further investigation, one might stop, if no additional alerting or confirming indicator were found. Or one might decide to explore the situation even further; perhaps even to hospitalize, for example. "In each case, the decision is not what to do for all time, but rather what to do next, for the near future (Pokorny, 1983) . . . there is continuing opportunity for feedback, and thus for modification of risk assessment and intervention. (p. 53)

From the standpoint of any potential malpractice action, the most crucial element in the formulation of clinical judgment is that the psychologist's professional behavior not significantly deviate from what is usual and customary for the care of patients with these particular signs and symptoms. That is to say, the psychologist will have demonstrated the behavior of a reasonable and prudent practitioner and not made any significant omissions in assessment, as well as have taken appropriate precautions to minimize the risk of a patient suicide (Berman & Cohen-Sandler, 1982, 1983). The importance of thorough documentation cannot be overstated.

Patient suicide is not a tragedy that exclusively confronts the mental health professions, but "the incidence of its occurrence is such a frequent issue, both professionally and sometimes legally, that it demands special consideration" from mental health professionals (J. Smith, 1986, p. 62). Authorities who investigate suicidal phenomena have not reached consensus on the key risk factors, whether short-term and long-term, that distinguish suicide completers. Ironically, for the practicing clinician, this lack of consensus offers some protection from a legal perspective. Courts and juries have often held that when it comes to suicide, there is no single correct or perfect solution in the management of the suicidal patient. Rather, both courts and juries tend to judge the clinician's actions in comparison with what seemed reasonable in compliance with the accepted standards of their profession (Simon, 1988).

Finally, the psychologist must understand that the final decision as to suicide risk is an intuitive judgment, that "we are obliged to accept that no matter how much information is gathered, sooner or later all the data must be weighed together and an intuitive estimate of risk recorded. That it is only an educated guess does not diminish its importance or its value as a consideration in management and treatment planning" (Motto, 1989, p. 2).

NOTES

1. Portions of this chapter are adapted from B. Bongar, *The suicidal patient: Clinical and legal standards of care* (2nd ed.). Copyright: American Psychological Association.

REFERENCES

Abram, H. S., Moore, G. I., & Westervelt, F. B. (1971). Suicidal behavior in chronic dialysis patients. *American Journal of Psychiatry, 127*, 1199.

Alcohol, Drug Abuse, and Mental Health Administration (1989). *Report of the Secretary's Task Force*

on Youth Suicide (Vols. 1–4). DHSS Publication No. ADM 89-1621-1624). Washington, DC: U.S. Government Printing Office.

Arana, G.W., & Hyman, S. (1989). Biological contributions to suicide. In D. G. Jacobs & H. N. Brown (Eds.), *Suicide: Understanding and responding: Harvard Medical School perspectives on suicide* (pp. 73–86). Madison, CT: International Universities Press.

Asberg, M., Bertillson L., & Martensson, B. (1984). CSF monoamine metabolites, depression, and suicide. *Advances in Biochemical Psychopharmacology* 39, 87–97.

Asnis, G. M., Friedman, T. A., Sanderson, W. C., Kaplan, M. L., van Praag, H. M., & Harkavy-Friedman, J. M. (1993). Suicidal behaviors in adult psychiatric outpatients. I: Description and prevalence. *American Journal of Psychiatry, 150*(1), 108–112.

Bassuk, E. L. (1982). General principles of assessment. In E. L. Bassuk, S. C. Schoonover, & A. D. Gill (Eds.), *Lifelines: Clinical perspectives on suicide* (pp. 17–46). New York: Plenum.

Beck, A. T. (1986). Hopelessness as a predictor of eventual suicide. In J. J. Mann & M. Stanley (Eds.), *Psychobiology* (pp. 90–96). New York: New York Academy of Sciences.

Beck, A. T. (1989). An interview with a depressed and suicidal patient. In D. Wedding & R. J. Corsini (Eds.), *Case studies in psychotherapy* (pp. 125–142). Itasca, IL: F. E. Peacock.

Beck, A. T., Brown, G., Berchick, R. J., Stewart, B. L., & Steer, R. A. (1990). Relation of hopelessness to ultimate suicide: A replication with psychiatric outpatients. *American Journal of Psychiatry, 147*(2), 190–195.

Beck, A. T., Brown, G. K., Steer, R. A., Dahlsgaard, K. K., & Grisham, J. R. (1999). Suicide ideation at its worst point: A predictor of eventual suicide in psychiatric outpatients. *Suicide & Life-Threatening Behavior, 29*(1), 1–9.

Beck, A. T., Kovacs, M., & Weissman, A. (1979). Assessment of suicidal intention: The scale for suicide ideation. *Journal of Consulting and Clinical Psychology, 47*, 343–352.

Beck, A. T., Schuyler, D., & Herman, I. (1974). Development of suicidal intent scales. In A. T. Beck, H. L. P. Resnik, & D. J. Lettieri (Eds.), *The prediction of suicide* (pp. 45–56). Bowie, MD: Charles Press.

Beck, A. T., & Steer, R. A. (1987). *Beck Depression Inventory: Manual*. San Antonio, TX: The Psychological Corporation.

Beck, A. T., & Steer, R. A. (1988). *Beck Hopelessness Scale: Manual*. San Antonio, TX: The Psychological Corporation.

Beck, A. T., Steer, R. A., Kovacs, M., & Garrison, B. (1985). Hopelessness and eventual suicide: A 10 year prospective study of patients hospitalized with suicidal ideation. *American Journal of Psychiatry, 142*, 559–563.

Beck, A. T., Steer, R. A., Ranieri, W. F. (1988). Scale for suicide ideation: Psychometric properties of a self-report version. *Journal of Clinical Psychology, 44*(4), 500–505.

Beck, A. T., Weissman, A., Lester, D., & Trexler, L. (1974). The measurement of pessimism: The Hopelessness Scale. *Journal of Consulting and Clinical Psychology, 42*(1), 861–865.

Berman, A. L., & Cohen-Sandler, R. (1982). Suicide and the standard of care: Optimal vs. acceptable. *Suicide and Life Threatening Behavior, 12*(2), 114–122.

Berman, A. L., & Cohen-Sandler, R. (1983). Suicide and malpractice: Expert testimony and the standard of care. *Professional Psychology: Research and Practice, 14*(1), 6–19.

Berman, A. L., & Jobes, D. (1991). *Adolescent suicide*. Washington, DC: American Psychological Association.

Beutler, L. E. (1989). Differential treatment selection: The role of diagnosis in psychotherapy. *Psychotherapy, 26*(3), 271–281.

Black, D. W., & Winokur, G. (1990). Suicide and psychiatric diagnosis. In S. J. Blumenthal & D. J. Kupfer (Eds.), *Suicide over the life cycle: Risk factors, assessment, and treatment of suicidal patients* (pp. 135–154). Washington, DC: American Psychiatric Press.

Bongar, B. (1991). *The suicidal patient: Clinical and legal standards of care*. Washington, DC: American Psychological Association.

Bongar, B. (1992a). *Suicide: Guidelines for assessment, management, and treatment*. New York: Oxford University Press.

Bongar, B. (1992b). The ethical issue of competence in working with the suicidal patient. *Ethics and Behavior, 2*(2), 75–89.

Bongar, B. (1993). A rejoinder to Bongar revisited: Tort reform as an element in system change. *The Register Report: National Register of Health Service Providers in Psychology, 19*(2), 15–16.

Bongar, B. (1998). Introduction. In B. Bongar, A. L. Berman, R. W. Maris, M. M. Silverman, E. A. Harris, & W. L. Packman (Eds.), *Risk management with suicidal patients* (pp. 1–3). New York: Guilford.

Bongar, B. (2000). Suicide. In C. W. Edward & C. B. Nemeroff (Eds.), *The Corsini encyclopedia of psychology and the behavioral sciences* (3rd ed.). New York: Wiley.

Bongar, B., Greaney, S., & Peruzzi, N. (1998). Risk management with the suicidal patient. In P. Kleespies (Ed.), *Emergencies in mental health practice: Evaluation and management* (pp. 199–216). New York: Guilford.

Bongar, B., Maris, R. W., Berman, A. L., & Litman, R. E. (1992). Outpatient standards of care and the suicidal patient. *Suicide and Life-Threatening Behavior, 22*(4), 453–478.

Bongar, B., Maris, R. W., Berman, A. L, Litman, R. E., & Silverman, M. M. (1993, Summer). Inpatient standards of care and the suicidal patient: General clinical formulations and legal considerations. *Suicide and Life-Threatening Behavior 23*(2).

Bongar, B., Peruzzi, N., & Greaney, S. (in press). Risk management and the suicidal patient. In Kleespies, P. (Ed.), *Handbook of psychological emergencies*.

Bongar, B., Peterson, L. G., Harris, E. A., & Aissis, J. (1989). Clinical and legal considerations in the management of suicidal patients. An integrative overview. *Journal of Integrative and Eclectic Psychotherapy*, 8, 53–67.

Boyd, J. (1983). The increasing rate of suicide by firearms. *New England Journal of Medicine, 308*, 872–898.

Brent, D. A., Kupfer, D. J., Bromet, E. J., & Dew, M. A. (1988). The assessment and treatment of patients at risk for suicide. In A. J. Frances & R. E. Hales (Eds.), *American Psychiatric Press Review of Psychiatry Vol. 7* (pp. 353–385). Washington, DC: American Psychiatric Press.

Brown, L., Ebert, M. H., Goyer, P. F., Jimerson, D., Klein, W. J., Bunney, W. E., & Goodwin, F. K. (1982). Aggression, suicide and serotonin: Relationship to CSF amine metabolites. *American Journal of Psychiatry, 139*(6), 741–746.

Bruno, G. (1995). *Suicide assessment practices among psychologists*. Unpublished doctoral dissertation, Pacific Graduate School of Psychology, Palo Alto, CA.

Buglass, D., & McCulloch, J. W. (1970). Further suicidal behaviour: the development and validation of predictive scales. *British Journal of Psychiatry, 116*(534), 483–491.

Butcher, J. N. (1989). *The Minnesota Report: Adult Clinical System MMPI-2*. Minneapolis: University of Minnesota Press.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI-2: Manual for administration and scoring*. Minneapolis: University of Minnesota Press.

Canapary, A. (1999). Suicidal risk factors with an alcoholic population. Unpublished doctoral dissertation, Pacific Graduate School of Psychology, Palo Alto, CA.

Card-Strong, L. (1998). *A national survey of psychologists' practice behaviors when diagnosing and treating major depressive disorder*. Unpublished doctoral dissertation, Pacific Graduate School of Psychology, Palo Alto, CA.

Chemtob, C. M., Bauer, G. B., Hamada, R. S., Pelowski, S. R., & Muraoka, M. Y. (1989). Patient suicide: Occupational hazard for psychologists and psychiatrists. *Professional Psychology: Research and Practice, 20*(5), 294–300.

Choca, J. P. (1992). *Interpretative guide to the Millon Clinical Multiaxial Inventory*. Washington, DC: American Psychological Association.

Clark, D. C. (1990). Panic disorder: Summary and editor's commentary. *Suicide Research Digest, IV*(1), 3–4.

Clark, D. C., & Fawcett, J. (1992). Review of empirical risk factors for evaluation of the suicidal patient. In B. Bongar (Ed.), *Suicide: Guidelines for assessment, management and treatment* (pp. 16–48). New York: Oxford University Press.

Clark, D. C., Young, M. A., Scheftner, W. A., Fawcett, J., & Fogg, L. (1987). A field test of Motto's Risk Estimator for Suicide. *American Journal of Psychiatry, 144*(7), 923–926.

Clopton, J. R. (1974). Suicidal risk via the Minnesota Multiphasic Personality Inventory (MMPI). In C. Neuringer (Ed.), *Psychological assessment of suicide risk* (pp. 118–133). Springfield, IL: Charles C. Thomas.

Cohen, E., Motto, J., & Seiden, R. (1966). An instrument for evaluating suicide potential. *American Journal of Psychiatry, 122*, 886–891.

Cole, J. K., & Magnussen, M. (1966). Where the action is. *Journal of Consulting and Clinical Psychology, 30*(1), 539–545.

Cottle, T. J. (1976). *Perceiving time*. New York: Wiley.

Craig, R. J., and Olson, R. E. (1990). MMPI characteristics of drug abusers with and without histories of suicide attempts. *Journal of Personality Assessment, 55*, 717–728.

Cull, J. G., & Gill, W. S. (1982). *Suicide Probability Scale (SPS) Manual*. Los Angeles: Western Psychological Services.

Cummings, N. A., & VandenBos, G. R. (1979). The general practice of psychology. *Professional Psychology: Research and Practice, 10*(1), 430–440.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom,

L. E. (1972). *An MMPI handbook* (Vol. 1). Minneapolis: University of Minnesota Press.

Eisenthal, S. (1974). Assessment of suicide risk using selected tests. In C. Neuringer (Ed.), *Psychological assessment of suicide risk* (pp. 134–149). Springfield, IL: Charles C. Thomas.

Ellis, T. E., Rudd, M. D., Rajab, M. H., & Wehrly, T. E. (1996). Cluster analysis of MCMI scores of suicidal psychiatric patients: Four personality profiles. *Journal of Clinical Psychology, 52*(4), 411–422.

Evans, G., & Farberow, N. L. (1988). *The encyclopedia of suicide*. New York: Facts on File.

Exner, J. (1978). *The Rorschach: A comprehensive system. Vol. 2. Current research and advanced interpretation*. New York: Wiley.

Exner, J. (1986). *The Rorschach: A comprehensive system: Vol. 1* (2nd ed). New York: Wiley.

Exner, J. E., and Weiner, I. B. (1982). *The Rorschach: A comprehensive system. Vol. 3. Assessment of children and adolescents*. New York: Wiley.

Eyman, J. R. (1987). *Unsuccessful psychotherapy with seriously suicidal borderline patients*. Paper presented at the joint meeting of the American Association of Suicidology and the International Association of Suicidology, San Francisco.

Eyman, J. R., & Eyman, S. K. (1991). Personality assessment in suicide prediction. *Suicide and Life-Threatening Behavior, 21*(1), 37–55.

Farberow, N. L., Helig, S., & Litman, R. (1968). *Techniques in crisis intervention: A training manual*. Los Angeles: Suicide Prevention Center.

Fawcett, J. (1988, May 7). Interventions against suicide. In D. G. Jacobs and J. Fawcett (Chairs), *Suicide and the psychiatrist: Clinical challenges*. Symposium presented by the Suicide Education Institute of Boston, in collaboration with The Center of Suicide Research and Prevention, at the American Psychiatric Association Annual Meeting, Montreal.

Fawcett, J., Scheftner, W., Clark, D., Hedeker, D., Gibbons, R., & Coryell, W. (1987). Clinical predictors of suicide in patients with major affective disorders: A controlled prospective study. *American Journal of Psychiatry, 144*(1), 35–40.

Firestone, R. W. & Firestone, L. A. (1996). Firestone Assessment of Self-Destructive Thoughts Manual. San Antonio: Harcourt Brace & Company.

Firestone, R. W., & Seiden, R. H. (1990). Suicide and the continuum of self-destructive behavior. *Journal of American College Health, 38*(5), 207–213.

Frank, G. (1994). On the prediction of suicide from the Rorschach. *Psychological Reports, 74*(3 Pt. 1), 787–794.

Frederick, C. J. (1978). Current trends in suicidal behavior. *American Journal of Psychotherapy, 32*(2), 172–200.

Fremouw, W. J., de Perczel, M., & Ellis, T. E. (1990). *Suicide risk: Assessment and response guidelines*. New York: Pergamon Press.

Glassmire, D. M., Stolberg, R. A., Greene, R. L., & Bongar, B. (2001) The utility of suicide items for assessing suicide potential: Development of a suicide potential scale. *Assessment, 8*(3), 281–290.

Goldring, N., & Fieve, R. R. (1984). Attempted suicide in manic-depressive disorder. *American Journal of Psychotherapy, 38*(3), 373–383.

Greaney, S. (1995). *Psychologists behavior and attitudes when working with the non-hospitalized suicidal patient*. Unpublished doctoral dissertation, Pacific Graduate School of Psychology, Palo Alto, CA.

Greene, R. L. (2000). *The MMPI-2: An interpretive manual* (2nd ed.). Boston: Allyn & Bacon.

Gunderson, J. G., Kolb, J. E., & Austin, V. (1981). The diagnostic interview for borderline patients. *American Journal of Psychiatry, 138*, 896–903.

Gutheil, T. G., & Appelbaum, P. S. (1982). *Clinical handbook of psychiatry and the law*. New York: McGraw-Hill.

Hendren, R. L. (1990). Assessment and interviewing strategies for suicidal patients over the life cycle. In S. J. Blumenthal & D. J. Kupfer (Eds.), *Suicide over the life cycle: Risk factors, assessment, and treatment of suicidal patients* (pp. 235–252). Washington, DC: American Psychiatric Press.

Hirschfeld, R. M. A., & Davidson, L. (1988). Risk factors for suicide. In A. J. Frances & R. E. Hales (Eds.), *American Psychiatric Press Review of Psychiatry* (Vol. 7) (pp. 307–333). Washington, DC: American Psychiatric Press.

Hirschfeld, R. M. A., & Russell, J. M. (1997). Assessment and treatment of suicidal patients. *New England Journal of Medicine, 333*(13), 910–915.

Inwald, R. E., Brobst, K. E., & Morrissey, R. F. (1987). *Hillson Adolescent Profile Manual*. Kew Gardens, NY: Hillson Research.

Jacobs, D. G. (1989a). Evaluation and care of suicidal behavior in emergency settings. In D. G. Jacobs & H. N. Brown (Eds.), *Suicide: Understanding and responding: Harvard Medical School perspectives on suicide* (pp. 363–377). Madison, CT: International Universities Press.

Jacobs, D. G. (1989b). Psychotherapy with suicidal patients: The empathic method. In D. G. Jacobs & H. N. Brown (Eds.), Suicide: *Understanding and responding: Harvard Medical School perspectives on suicide* (pp. 329–342). Madison, CT: International Universities Press.

Jobes, D. A., Eyman, J. R., & Yufit, R. I. (1990). *Suicide risk assessment survey*. Paper presented at the annual meeting of the American Association of Suicidology, New Orleans.

Joiner, T. E. Jr., Rudd, M. D., & Rajab, M. H. (1997). The Modified Scale for Suicidal Ideation: factors of suicidality and their relation to clinical and diagnostic variables. *Journal of Abnormal Psychology, 106*(2), 260–265.

Joiner, T. E., Walker, R. L., Rudd, M. D., & Jobes, D. A. (1999). Scientizing and routinizing the assessment of suicidality in outpatient practice. *Professional Psychology: Research and Practice, 30*(5), 447–453.

Kaplan, M. L., Asnis, G. M., Sanderson, W. C., & Keswani, L. (1994). Suicide assessment: Clinical interview vs. self-report. *Journal of Clinical Psychology, 50*(2), 294–298.

Kreitman, N. (Ed.). (1977). *Parasuicide*. New York: Wiley.

Kreitman, N. (1986). The clinical assessment and management of the suicidal patient. In A. Roy (Ed.), *Suicide* (pp. 181–195). Baltimore: Williams & Wilkins.

Lazarus, A. A. (1981). *The practice of multimodal therapy: Systematic, comprehensive and effective psychotherapy*. New York: McGraw-Hill.

Lesse, S. (1989). The range of therapies with severely depressed suicidal patients. In S. Lesse (Ed.), *What we know about suicidal behavior and how to treat it* (pp. 193–217). Northvale, NJ: Jason Aronson.

Lewinsohn, P. M., Garrison, C. Z., Langhinrichsen, J., & Marsteller, F. (1989). *The assessment of suicidal behavior in adolescents: A review of scales suitable for epidemiological and clinical research* (Report prepared under Contract Nos. 316774, 316776). Rockville, MD: Child and Adolescent Disorders Research Branch DCR, National Institute of Mental Health.

Lifton, R. J., Kato, S., & Reich, M. R. (1979). *Six lives/six deaths*. New Haven: Yale University Press.

Linehan, M. M. (1985). The reasons for living inventory. In P. Keller & L. Ritt (Eds.), *Innovations in clinical practice* (pp. 321–330). Sarasota, FL: Professional Resource Exchange.

Linehan, M. M. (1987). Dialectical behavior therapy: A cognitive behavioral approach to parasuicide. *Journal of Personality Disorders, 1*(4), 328–333.

Linehan, M. M. (1999). Standard protocol for assessing and treating suicidal behaviors for patients in treatment. In D. G. Jacobs et al. (Eds.), *The Harvard Medical School guide to suicide assessment and intervention* (pp. 146–187). San Francisco: Jossey-Bass.

Linehan, M. M., Rizvi, S. L., Welch, S. S., & Page, B. (2000). Suicide and personality disorders. In K. Hawton & K. van Heeringen (Eds.), *International handbook of suicide and attempted suicide*. Sussex, England: John Wiley & Sons.

Litman, R. E. (1975). The assessment of suicidality. In R. Pasnau (Ed.), *Consultation-liasion psychiatry*. New York: Grune and Stratton.

Litman, R. E. (1989). Suicides: What do they have in mind? In D. G. Jacobs & H. N. Brown (Eds.), *Suicide: Understanding and responding: Harvard Medical School perspectives on suicide* (pp. 143–154). Madison, CT: International Universities Press.

Mack, J. E., & Hickler, H. (1981). *Vivienne: The life and suicide of an adolescent girl*. Boston: Little Brown.

Mahrer, J. (1993). *The use of "no-suicide contracts" and agreements with suicidal patients. Dissertation Abstracts International, 54*(6B), 3345.

Maltsberger, J. T. (1988). Suicide danger: Clinical estimation and decision. *Suicide and Life-Threatening Behavior, 18*(1), 47–54.

Mann, J. J., Oquendo, M., Underwood, M. D., & Arango, V. (1999). The neurobiology of suicide risk: A review for the clinician. *Journal of Clinical Psychiatry, 60*(Suppl. 2), 7–11.

Mann J. J., & Stanley, M. (1988a). Foreword to Section III. Suicide. In A. J. Frances & R. E. Hales (Eds.), *American Psychiatric Press Review of Psychiatry* (Vol. 7) (pp. 287–288). Washington, DC: American Psychiatric Press.

Mann J. J., & Stanley, M. (1988b). Afterword to Section III: Suicide. In A. J. Frances & R. E. Hales (Eds.), *American Psychiatric Press Review of Psychiatry Vol. 7* (pp. 422–426). Washington, DC: American Psychiatric Press.

Maris, R. W. (1981). *Pathways to suicide: A survey of self-destructive behaviors*. Baltimore: Johns Hopkins University Press.

Maris, R. W. (1986). Preface to the special issue: Biology of suicide. *Suicide and Life-Threatening Behavior 16*(2), v–viii.

Maris, R. W. (1988). Preface: Overview and discussion. In R. W. Maris (Ed.), *Understanding and preventing suicide: Plenary papers of the first combined meeting of the AAS and IASP* (pp. vii–xxiii). New York: Guilford Press.

Maris, R. W. (1989). Preface: Strategies for studying suicide and suicidal behavior. *Suicide and Life-Threatening Behavior, 19*(1), ix–x.

Maris, R. W., Berman, A. L., Maltsberger, J. T., & Yufit, R. I. (Eds.) (1992). *Assessment and prediction of suicide*. New York: Guilford Press.

McCann, J. T., & Gergelis, R. E. (1990). Utility of the MCMI-II in assessing suicide risk. *Journal of Clinical Psychology, 46*(6), 764–770.

McCann, J. T., & Suess, J. F. (1988). Clinical

applications of the MCMI: The 1-2-3-8 code type. *Journal of Clinical Psychology, 44*(2), 181–186.

McEvoy, T. L. (1974). Suicidal risk via the Thematic Apperception Test. In C. Neuringer (Ed.), *Psychological assessment of suicide risk* (pp. 3–17). Springfield, IL: Charles C. Thomas.

McIntosh, J. L., Santos, J. F., Hubbard, R. W., & Overholser, J. C. (1994). *Elder suicide: Research, theory and treatment*. Washington, DC: American Psychological Association.

McIntosh, J., et al. (forthcoming). *Suicide in the elderly*. Washington, DC: American Psychological Association.

Meltzer, H. Y. (1998). Suicide in schizophrenia: Risk factors and clozapine treatment. *Journal of Clinical Psychiatry, 59*(Suppl. 3), 15–20.

Meyer, R. G. (1989). *The clinicians handbook: The psychopathology of adulthood and adolescence* (2nd ed.). Boston: Allyn & Bacon.

Meyer, R. G. (1993). *The clinician's handbook: Integrated diagnostics, assessment, and intervention in adult and adolescent psychopathology* (3rd ed.). Boston: Allyn & Bacon.

Meyer, R. G., & Deitsch, S. E. (1996). *The clinician's handbook: Integrated diagnostics, assessment, and intervention in adult and adolescent psychopathology* (4th ed.). Boston: Allyn & Bacon.

Miskimins, R. W., DeCook, R., Wilson, L. T., & Maley, R. F. (1967). Prediction of suicide in a psychiatric hospital. *Journal of Clinical Psycholgy, 23*, 296–301.

Monahan, J. (1981). The clinical prediction of violent behavior (DHSS Publication No. ADM 81-921). Washington, DC: National Institute of Mental Health.

Motto, J. A. (1977). The estimation of suicide risk by use of clinical models. *Suicide and Life-Threatening Behavior, 7*(4), 236–245.

Motto, J. A. (1979). The psychopathology of suicide: A clinical approach. *American Journal of Psychiatry, 136*(4-B), 516–520.

Motto, J. A. (1986). Clinical considerations of biological correlates of suicide. *Suicide and Life-Threatening Behavior, 16*(2), 1–20.

Motto, J. A. (1989). Problems in suicide risk assessment. In D. G. Jacobs & H. N. Brown (Eds.), *Suicide: Understanding and responding: Harvard Medical School perspectives on suicide* (pp. 129–142). Madison, CT: International Universities Press.

Motto, J. A. (1991). An integrated approach to estimating suicide risk. *Suicide and Life-Threatening Behavior, 21*(1), 74–89.

Motto, J. A., Heilbron, D. C., & Juster, R. P. (1985). Development of a clinical instrument to estimate suicide risk. *American Journal of Psychiatry, 142*(6), 680–686.

Murphy, G. E. (1986). The physician's role in suicide prevention. In A. Roy (Ed.), *Suicide* (pp. 171–179). Baltimore: Williams & Wilkins.

Murphy, G. E. (1988a). The prediction of suicide. In S. Lesse (Ed.), *What we know about suicidal behavior and how to treat it* (pp. 47–58). Northvale, NJ: Jason Aronson.

Murphy, G. E. (1988b). Prevention of suicide. In A. J. Frances & R. E. Hales (Eds.), *American Psychiatric Press Review of Psychiatry Vol. 7* (pp. 403–421). Washington, DC: American Psychiatric Press.

Murray, H. (1938). *Explorations in personality*. New York: Oxford University Press.

Nemiah, J. C. (1982). Foreword. In F. L. Bassuk, S. C. Schoonover, & A. D. Gill (Eds.), *Lifelines: Clinical perspectives on suicide*. New York: Plenum.

Neuringer, C. (1974a). Problems in assessing suicide risk. In C. Neuringer (Ed.), *Psychological assessment of suicide risk* (pp. 3–17). Springfield, IL: Charles C. Thomas.

Neuringer, C. (1974b). Rorschach inkblot test assessment of suicidal risk. In C. Neuringer (Ed.), *Psychological assessment of suicide risk* (pp. 74–94). Springfield, IL: Charles C. Thomas.

Norcross, J. (1986). Training integrative/electic psychotherapists. *Journal of Integrative and Eclectic Psychotherapy, 5*(1), 71–94.

Osborne, D. (1985). The MMPI in psychiatric practice. *Psychiatric Annals, 15*(9), 542–545.

Pallis, D. J., Barraclough, B., Levey, A., Jenkins, J., & Sainsbury, P. (1982). Estimating suicide risk among attempted suicides: The development of new clinical scales. *British Journal of Psychiatry, 141*, 37–44.

Pallis, D. J., Gibbons, J. S., & Pierce, D. W. (1984). Estimating suicide risk among attempted suicides: II. Efficiency of predictive scales after the attempt. *British Journal of Psychiatry, 144*, 139–148.

Perr, I. N. (1985). Suicide litigation and risk management: A review of 32 cases. *Bulletin of American Academy of Psychiatry and the Law, 13*(3), 209–219.

Perr, I. N. (1988). The practice of psychiatry and suicide litigation. *New Developments in Mental Health Law, 8*(1), 4–19.

Peruzzi, N., & Bongar, B. (1999). Assessing risk for completed suicide in patients with major depression: Psychologists' views of critical factors. *Professional Psychology: Research & Practice, 30*(6), 576–580.

Peterson, L. G., & Bongar, B. (1989). The suicidal patient. In A. Lazare (Ed.), *Outpatient psychiatry: Diagnosis and treatment* (2nd ed., pp. 569–584). Baltimore: Williams & Wilkins.

Peterson, L. G., Bongar, B., & Netoski, M. (1989). Regional use of violent suicidal methods: An analysis of suicide in Houston, Texas. *American Journal of Emergency Medicine, 7*(2), 21–27.

Piotrowski, C., Sherry, D., & Keller, J. W. (1985). Psychodiagnostic test usage: A survey of the Society for Personality Assessment. *Journal of Personality Assessment, 49*(1), 115–119.

Poeldinger, W. J., Gehring, A., & Blaser, P. (1973). Suicide risk and MMPI scores, especially as related to anxiety and depression. *Suicide and Life-Threatening Behavior, 3*(2), 147–153.

Pokorny, A. D. (1964). A follow-up study of 618 suicidal patients. *American Journal of Psychiatry, 122*, 1109–1116.

Pokorny, A. D. (1983). Prediction of suicide in psychiatric patients. *Archives of General Psychiatry, 40*, 249–257.

Pokorny, A. D. (1993). Suicide prediction revisited. *Suicide and Life-Threatening Behavior, 23*(1), 1–10.

Pope, K. (1989, January). Malpractice suits, licensing disciplinary actions, and ethics cases. *Independent Practitioner*, pp. 22–26.

Quinnett, P., & Bratcher, K. (1996). Question Persuade Refer Treat: Suicide risk management inventory user's manual. Spokane, WA: Greentree Behavioral Health, a division of Spokane Mental Health.

Rachlin, S. (1984). Double jeopardy: Suicide and malpractice. *General Hospital Psychiatry, 6*, 302–307.

Reid, W. H., Mason, M., & Hogan, T. (1998). Suicide prevention effects associated with clozapine therapy in schizophrenia and schizoaffective disorder. *Psychiatric Service 49*(8), 1029–1033.

Reynolds, W. M. (1987). *Suicidal Ideation Questionnaire*. Odessa, FL: Psychological Assessment Resources.

Reynolds, W. M. (1988). Suicidal Ideation Questionnaire: Professional manual. Odessa, FL: Psychological Assessment Resources.

Reynolds, W. M. (1991). Adult Suicidal Ideation Questionnaire: Professional manual. Odessa, FL: Psychological Assessment Resources.

Reynolds, W. M., & Mazza, J. J. (1992). Suicidal Behavior History Form: Clinician's guide. Odessa, FL: Psychological Assessment Resources.

Rich, C. L., & Runeson, B. S. (1992). Similarities in diagnostic comorbidity between suicide among young people in Sweden and the United States. *Acta Psychiatrica Scandinavica, 86*(5), 335–339.

Richman, J. (1986). *Family therapy for suicidal people*. New York: Springer.

Richman, J., & Eyman, J. R. (1991). Psychotherapy of suicide: Individual, group and family approaches. In D. Lester (Ed.), *Understanding suicide: The state of the art*. Philadelphia: Charles Press.

Robins, E. (1985). Suicide. In H. I. Kaplan & B. J. Sadock (Eds.), *Comprehensive textbook of psychiatry: Vol. IV* (pp. 311–1315). Baltimore: Williams & Wilkins.

Robins, E. (1986). Completed suicide. In A. Roy (Ed.), *Suicide* (pp. 123–133). Baltimore: Williams & Wilkins.

Robins, L. N., & Helzer, J. E. (1986). Diagnosis and clinical assessment: The current state of psychiatric diagnosis. *Annual Review of Psychology*, 37, 409–432.

Rosenberg, J. I. (1997) Expertise research and clinical practice: A suicide assessment and intervention training model. *Educational Psychology Review*, 9(3), 279–296.

Roy, A. (1986). Preface. In A. Roy (Ed.), *Suicide* (p. vii). Baltimore: Williams & Wilkins.

Roy, A., Nielsen, D., Rylander, G., Sarchiapone, M., & Segal, N. (1999). Genetics of suicide in depression. *Journal of Clinical Psychiatry, 60*(Suppl. 2), 12–17.

Schein, H. M. (1976). Obstacles in the education of psychiatric residents. *Omega*, 7, 75–82.

Sepaher, I., Bongar, B., & Greene, R. L. (1999). Codetype base rates for the "I Mean Business" suicide items on the MMPI-2. *Journal of Clinical Psychology*, 55(9), 1167–1173.

Shneidman E. S. (1984). Aphorisms of suicide and some implications for psychotherapy. *American Journal of Psychotherapy*, 38(3), 319–328.

Shneidman, E. S. (1985). *Definition of suicide*. New York: Wiley.

Shneidman, E. S. (1986a). Some essentials of suicide and some implications for response. In A. Roy (Ed.), *Suicide* (pp. 1–16). Baltimore: Williams & Wilkins.

Shneidman, E. S. (1986b). Suicidal logic. In W. S. Sahakian, B. J. Sahakian, & P. L. Sahakian-Nunn (Eds.), *Psychopathology today: The current status of abnormal psychology* (3rd ed., pp. 267–281). Itasca, IL: Peacock Press.

Shneidman, E. S. (1987). A psychological approach to suicide. In G. R. VandenBos & B. K. Bryant (Eds.), *Cataclysms, crises, and catastrophes: Psychology in action* (pp. 147–183). Washington, DC: American Psychological Association.

Shneidman, E. S. (1989). Overview: A multidimensional approach to suicide. In D. G. Jacobs & H.

N. Brown (Eds.), *Suicide: Understanding and responding: Harvard Medical School perspectives on suicide* (pp. 1–30). Madison, CT: International Universities Press.

Silberg, J. L., & Armstrong, J. G. (1992). The Rorschach test for predicting suicide among depressed adolescent inpatients. *Journal of Personality Assessment, 59*(2), 290–303.

Simon, R. I. (1987). *Clinical psychiatry and the law*. Washington, DC: American Psychiatric Press.

Simon, R. I. (1988). *Concise guide to clinical psychiatry and the law*. Washington, DC: American Psychiatric Press.

Simon, R. I. (1992). *Concise guide to psychiatry and the law for clinicians*. Washington, DC: American Psychiatric Press.

Smith, J. (1986). *Medical malpractice psychiatric care*. Colorado Springs, CO: Shepards/McGraw-Hill.

Smith, K. (1985). Suicide assessment: An ego vulnerabilities approach. *Bulletin of the Menninger Clinic, 48*(5), 489–499.

Smith, K., & Eyman, J. R. (1988). Ego structure and object differentiation in suicidal patients. In H. D. Lerner, & P. M. Lerner (Eds.), *Primitive mental states and the Rorschach* (pp. 175–202). Madison, CT: International Universities Press.

Stanley, M., & Mann, J. J. (1988). Biological factors associated with suicide. In A. J. Frances & R. E. Hales (Eds.), *American Psychiatric Press Review of Psychiatry Vol. 7* (pp. 334–352). Washington, DC: American Psychiatric Press.

Steer, R. A., & Beck, A. T. (1988). Use of the Beck Depression Inventory, Hopelessness Scale, Scale for Suicide Ideation, and Suicidal Intent Scale with adolescents. *Advances in Adolescent Mental Health, 3*, 219–231.

Stolberg, R. A., Glassmire, D. M., Ricci, C. M., Green, R. L., & Bongar, B. (1999, April 17). *Using the MMPI-2 to increase the accuracy of identifying suicidal clients*. Poster presented at the 34th Annual Symposium on Recent Developments in the Use of the MMPI-2 & MMPI-A. Huntington Beach, CA.

Swartz, M. (1988). Malpractice liability for suicide: Clinical, legal and rhetorical dimensions. *The Psychiatric Forum, 14*(1), 45–53.

Swenson, E. V. (1986). Legal liability for a patient's suicide. *The Journal of Psychiatry & Law, 14*(3), 409–434.

Swiercinsky, D. (1985). *Testing adults*. Kansas City, MO: Test Corporation of America.

VandeCreek, L., & Knapp, S. (1983). Malpractice risks with suicidal patients. *Psychotherapy: Theory, Research and Practice, 20*(3), 274–280.

VandeCreek, L., & Knapp, S. (1989). *Tarasoff and beyond: Legal and clinical considerations in the treatment of life-endangering patients*. Sarasota, FL: Professional Resource Exchange.

Van Praag, H. M. (1986). Affective disorders and agression disorders: Evidence for a common biological mechanism. *Suicide and Life-Threatening Behavior 16*(2), 21–47.

Weissman, M. M., Klerman, G. L. Markowitz, J. S., & Ouelette, R. (1989). Suicidal ideation and suicide attempts in panic disorder and attacks. *New England Journal of Medicine, 321*, 1209–1214.

Weissman, A. D., & Worden, J. W. (1972). Risk-rescue in suicide assessment. *Archives of General Psychiatry, 26*, 553–560.

Weissman, A. D., & Worden, J. W. (1974). Risk-rescue in suicide assessment. In A. T. Beck, H. L. Resnik, & D. L. Lettieri (Eds.), *The prediction of suicide*. Bowies, MD: Charles Press.

Widiger, T., & Francis, A. (1987). Interviews and inventories for the measurement of personality disorders. *Clinical Psychology Review, 7*(1), 49–75.

Young, M. A., Fogg, L. F., Scheftner, W., Fawcett, J., Akiskal, H., & Maser, J. (1996). Stable trait components of hopelessness: Baseline and sensitivity to depression. *Journal of Abnormal Psychology, 105*, 155–165.

Yufit, R. I. (1988). *Manual of procedures—Assessing suicide potential: Suicide assessment team*. Unpublished manual. (Available from Robert I. Yufit, Ph.D., Department of Psychiatry and Behavioral Sciences, Division of Clinical Psychology, Northwestern University Medical School, Chicago, IL).

Yufit, R. I. (1989). Developing a suicide screening instrument for adolescents and young adults. In Alcohol, Drug Abuse, and Mental Health Administration *Report of the Secretary's Task Force on Youth Suicide* (DHSS Publication No. ADM 89-1621-1624). Washington, DC: U.S. Government Printing Office.

Yufit, R. I., & Benzies, B. (1973). Assessing suicide potential by time perpective. *Suicide and Life-Threatening Behavior, 3*(4), 270–282.

Yufit, R. I., & Benzies, B. (1979). *Preliminary Manual Time Questionnaire: Assessing suicide potential*. Palo Alto, CA: Consulting Psychologists Press.

Zubin, J. (1974). Observations on nosological issues in the classification of suicidal behavior. In A. T. Beck, H. L. Resnik, & D. L. Lettieri (Eds.). *The prediction of suicide* (pp. 3–25). Bowies, MD: Charles Press.

Zung, W. W. K. (1974). Index of Potential Suicide (IPS). In A. T. Beck, H. L. Resnik, & D. L. Lettieri (Eds.), *The prediction of suicide* (pp. 221–249). Bowies, MD: Charles Press.

25
Assessing Personality Disorders

Thomas A. Widiger
Linda Anne Coker

Personality disorders have been placed on a separate diagnostic axis (Axis II) in the modern editions of the American Psychiatric Association's (APA) *Diagnostic and Statistical Manual of Mental Disorders* (DSM). Most other mental disorders are placed on Axis I (APA, 2000). This unique distinction is to encourage the recognition of maladaptive personality traits in virtually every patient for whom a diagnosis might be provided, because personality traits will affect the occurrence, expression, course, and/or treatment of most Axis-I disorders (Frances, 1980). In addition, a substantial proportion of patients will be in treatment primarily because of a personality disorder. Personality disorders are among the most difficult of the mental disorders to treat, but they are not untreatable (Perry, Banon, & Ianni, 1999; Sanislow & McGlashan, 1998). Clinically significant and meaningful improvements to personality functioning are often obtained.

The assessment of personality disorders should then be of substantial importance to clinicians (Widiger, in press-a). Clinicians who are concerned primarily with the treatment of an Axis-I mental disorder should include a measure of personality disorder symptomatology if they intend to account fully for the variation in their patients' treatment responsivity, and clinicians who are concerned primarily with the treatment of a personality disorder are advised to include objective measures of personality disorder symptomatology to document empirically the effectiveness of their intervention.

Personality disorders, however, can be difficult to assess. One of the major innovations of DSM-III (APA, 1980) was the provision of relatively more specific and explicit diagnostic criterion sets that facilitated substantially the obtainment of more reliable clinical diagnoses. This innovation, however, has been problematic for the personality disorders. It is difficult, perhaps even impossible, to provide a brief list of specific diagnostic criteria for the broad and complex behavior patterns that constitute a personality disorder. As acknowledged in DSM-III-R, "for some disorders . . . , particularly the Personality Disorders, the criteria require much more inference on the part of the observer" (APA, 1987, p. xxiii). As a result, the only personality disorder to be diagnosed reliably in general clinical practice has been antisocial, and the validity of this diagnosis has been questioned precisely because of its emphasis on overt and behaviorally specific acts of criminality, irresponsibility, and delinquency (Hare, Hart, & Harpur, 1991).

Faraone, S. V., & Tsuang, M. T. (1993). Empirical determination of thresholds for case identification. Validation of the Personality Diagnostic Questionnaire-Revised. *Comprehensive Psychiatry, 34,* 384–391.

Zanarini, M. C., Frankenburg, F. R., Chauncey, D. L., & Gunderson, J. G. (1987). The Diagnostic Interview for Personality Disorders: Interrater and test-retest reliability. *Comprehensive Psychiatry, 28,* 467–480.

Zanarini, M. C., Gunderson, J. G., Frankenburg, F. R., & Chauncey, D. L. (1989). The Revised Diagnostic Interview for Borderlines: Discriminating BPD from other Axis II disorders. *Journal of Personality Disorders, 3,* 10–18.

Zarrella, K. L., Schuerger, J. M., & Ritz, G. H. (1990). Estimation of MCMI DSM-III Axis II constructs from MMPI scales and subscales. *Journal of Personality Assessment, 55,* 195–201.

Zimmerman, M. (1994). Diagnosing personality disorders. A review of issues and research methods. *Archives of General Psychiatry, 51,* 225–245.

Zimmerman, M., & Coryell, W. H. (1990). Diagnosing personality disorders in the community. A comparison of self-report and interview measures. *Archives of General Psychiatry, 47,* 527–531.

Zimmerman, M., & Mattia, J. I. (1999a). Differences between clinical and research practices in diagnosing borderline personality disorder. *American Journal of Psychiatry, 156,* 1570–1574.

Zimmerman, M., & Mattia, J. I. (1999b). Psychiatric diagnosis in clinical practice: Is comorbidity being missed? *Comprehensive Psychiatry, 40,* 182–191.

Zimmerman, M., Pfohl, B., Coryell, W., Stangl, D., & Corenthal, C. (1988). Diagnosing personality disorders in depressed patients. A comparison of patient and informant interviews. *Archives of General Psychiatry, 45,* 733–737.

26

Assessing the Risk of Aggression and Violence

Edwin I. Megargee

No clinical task is more challenging than assessing the risk of aggressive or violent behavior, and none can have more far-reaching consequences for all concerned. A person labeled "dangerous" can be served with a restraining order; lose parental rights in a custody dispute; be denied bail, probation, or parole; and, in certain circumstances, even be subjected to indefinite preventive, detention[1] (Stone, 1975). On the other hand, casualties can result from failure to predict future violence, as the writer discovered when he was pinned down for 90 minutes by Charles Whitman, a University of Texas student whose earlier confession of urges to go to the Texas Tower and shoot everyone in sight were not taken seriously by a Student Health Service psychiatrist; 38 people were shot, 14 fatally, before Whitman himself was slain.

As if such events were not enough to cause clinicians concern, lawsuits can result from the allegedly incorrect assessment of aggressive potential. In a series of cases beginning with *Tarasoff* v. *Regents of the University of California* in 1976, the courts have found that mental health workers who fail to appreciate the dangerousness of their clients and warn their potential victims may be liable for civil damages (Meyers, 1986; Stone, 1986).

It is not merely the gravity of the possible consequences that makes assessing the risk of aggression or violence so challenging. When appraising intelligence or diagnosing a thought disorder, clinicians can concentrate on the characteristics of their clients. However, by its very nature, aggressive behavior involves interpersonal interactions between two or more people, interactions that occur in a sociocultural context and that may involve constantly changing situational influences as the "scenario of violence" unfolds (Toch, 1969). Families, gangs, social groups, or even, in certain specialized circumstances, governments may be the subjects of psychologists' assessments of aggressive proclivities. For those venturing to predict violence, the task is further complicated by the low base rate for violence in most settings, which inflates the effects of even modest false-positive rates (Finn & Kamphuis, 1995; Meehl & Rosen, 1955; Megargee, 1976, 1981; Monahan, 1981).

The primary purpose of this chapter is to assist clinicians in understanding and assessing the factors that lead to aggressive behavior. After presenting a brief working definition of aggression, it discusses the context and purposes of such assessments, how they influence pro-

cedures, and some of the ethical and legal pitfalls that may be encountered. The next section provides a broad overview of data and methods that may be used in assessing aggressive behavior. This leads to a description of the "Algebra of aggression," a conceptual framework designed to help clinicians sort out and integrate relevant information and apply it to the problem of assessing the risk for dangerous behavior.

DEFINING AGGRESSION

Authorities often disagree on what constitutes aggression. Consequently, different theorists and researchers discussing "aggression" may be using incompatible semantic or operational definitions (Baron, 1977; A. H. Buss, 1961; Johnson, 1972). Is it aggression if a mosquito bites you on the arm? If you swat the mosquito? What if the mosquito you slap is on your spouse's cheek?

Recognizing that there are covert, indirect, unconscious, and passive forms of aggressive behavior, and that aggression occurs among other species, the focus of this chapter will be on overt, direct, intentional human aggression, that is, on *physical or verbal behavior that can cause people distress, pain, or injury, or damage their property or reputations.* Such behavior will be viewed as aggressive whether or not hurting the victim was the aggressor's primary intention and regardless of whether the behavior is classified as legal or illegal in the particular society in which it occurs.

DEFINING THE PURPOSE AND CONTEXT OF THE ASSESSMENT

The first step in any undertaking, whether it is building a birdhouse or planning an assessment, is to consider the purpose of the project and the setting or circumstances under which it will be carried out. Function dictates design, and, in the assessment of aggression, the context establishes the parameters within which an appraisal can be carried out.

Context of the Assessment

The physical setting and psychological context in which an assessment is conducted are always important, but this is especially so with regard to what has come to be known as "risk assessment," that is, the evaluation of the potential for individuals or groups to engage in aggression or violence in various situations. It is not uncommon for such appraisals to be conducted in jails, prisons, or closed psychiatric wards; sometimes a third party, such as an attorney, policy officer, or aide may be present. In some circumstances, the person being assessed is not even known to the assessor. As a consultant to the U.S. Secret Service, for example, the writer was occasionally asked to evaluate the threat posed by an anonymous, and hence unknown, letter writer.

In addition to the physical setting, the psychological and legal context may impose constraints on the assessment. The person being assessed often has a strong vested interest in the outcome; indeed in capital and death-row cases it may literally be a matter of life or death (Cunningham & Reidy, 1999). This can place the psychologist and the subject in an adversarial relationship. Clinicians doing assessments that include an appraisal of people's potential for engaging in aggressive behavior or violence owe it to themselves to become fully informed of the general legal and ethical issues involved, as well as the specific laws and policies in their settings[2] (American Association for Correctional Psychology, 2000; American Psychological Association, 1992; Bennett, Bryant, Vandenbos, & Greenwood, 1990; Brodsky, 1991; Heilbrun & Heilbrun, 1995; Monahan, 1972; Pope, Butcher, & Seelen, 1993). They must clearly understand "who is the client" (Monahan, 1972, 1996) and communicate this to the people being assessed at the outset so that subjects can make an informed decision regarding whether or not they will participate, and if so, how much they will communicate.

Because aggressive behavior may lead to some sort of litigation, psychologists doing such assessments should be meticulous about documenting their observations and keeping complete records. At the very least, they should be thoroughly familiar with their state's or province's requirements about preserving clinical records and follow them to the letter. In some situations, it may be advisable to have written documentation that the person being appraised understands all the implications of the assessment and has given informed consent.

For your own protection, consider videotaping interviews in certain assessment situations. If you evaluate formerly violent people being considered for release from secure settings, some of those you recommend for release will probably recidivate. When you are being sued by their victims or their estates, your erstwhile clients will be portrayed as raging beasts that any competent clinician would have kept caged for life. It may help if you can show a tape that demonstrates how well controlled, civilized, and reasonable these "monsters" were when you assessed them.

Purpose of the Appraisal

Most assessments of aggressive behavior fall into one of two general categories: "retrospective" and "prospective." In a *retrospective* or "diagnostic" *assessment*, psychologists are confronted with aggression that has already occurred and are asked to explain it. For example, criminal liability might depend on whether the victim provoked the attack or if the perpetrator was legally sane at the time. Retrospective evaluation is especially important in treatment planning. Obviously it makes a difference whether a brain tumor, a change in medication, an abusive relationship, sadism, or altruism caused the aggression.

Some assessment techniques work much better in retrospective than they do in prospective assessments. For example, when administered to people who have committed assaultive crimes, the Overcontrolled Hostility (O-H) scale for the MMPI-2 can help identify those whose violence paradoxically stems from excessive inhibitions against the overt expression of any sort of aggression (Megargee, Cook, & Mendelsohn, 1967). Apparently their inability to express anger causes their instigation to aggression to build up to the point that it overwhelms even their massive defenses, resulting in an extremely violent act (Megargee, 1966). Their treatment needs differ greatly from those of other types of violent offenders. However, this syndrome is so rare that the O-H scale is of limited utility in predicting violent behavior.

In *prospective or "prognostic" evaluations* or "risk assessments," psychologists are asked whether a person, group, or even a nation is likely to commit aggression in the future. If so, what form is the aggressive behavior likely to take? Against whom or what is the aggressive behavior likely to be directed? Under what conditions is such aggression most likely to occur? Is the likelihood greater or lower than the base rate? What circumstances or interventions are likely to increase or decrease the risk of such behavior? (Borum, Fein, Vossekuil, & Berglund, 1999; Heilbrun & Heilbrun, 1995). The answers to these questions might be used to help determine whether patients or prisoners are ready for release or if a child should be returned to the custody of a possibly abusive parent. As anyone who has ever bet on a sporting event knows, it is much more difficult to predict what will happen in the future than it is to explain what has happen in the past.

It is also important to distinguish between *generalized* and *individualized* risk assessment. In generalized risk assessment, the task is to identify groups of people who are more likely than most to be involved in aggression or violence. Such generalized assessments often take place in mental health facilities or correctional institutions as a guide to management planning or release decisions. For example, Bohn (1979) used Megargee's MMPI-based offender classification system (Megargee, 1977; Megargee & Bohn with Meyer & Sink, 1979) to identify groups of prisoners who were (a) more likely to prey on others and (b) more likely to be victimized. By assigning these groups to separate living quarters, the rate of serious assaults in a federal correctional institution was reduced 46%.

In individualized risk assessment the task is to determine the threat posed by a particular person. This may involve *targeted* risk assess-

ment, that is, the likelihood that a given individual will aggress against specific targets, such as spouses, co-workers, schoolmates, or public figures (Borum et al., 1999). For example, Secret Service agents must evaluate the likelihood that those who threaten the president of the United States will actually carry out their threats, and, as noted earlier, therapists are required to warn potential victims if they are in danger from clients (*Tarasoff* v. *Regents of the University of California*, 1976). Evaluating the risk of aggressive behavior in a particular person is a more challenging task than delineating a group of people whose potential for violence exceeds the base rates, especially in cases where the individual being assessed has no history of aggression or violence. The actuarial risk assessment devices that are often useful in generalized risk assessments may have little utility in individualized or targeted risk assessments (Borum et al., 1999).

Risk assessment is further complicated by the fact that for an act of aggression to occur, three basic elements are required:

1. A potential aggressor, that is, a person who is willing and able to carry out the aggressive act in question
2. A target or suitable victim who will be the object of the aggressive behavior
3. Opportunity for the aggressor to attack the victim

This goes considerably beyond personality assessment. It also requires assessing the potential perpetrator's ability to formulate and carry out a plan of attack, assessing his or her technical skills and access to weapons, predicting whether a suitable victim will be available, and finally, determining whether the would-be aggressor can obtain access to that victim (Cohen & Felson, 1979).

ASSESSING THE LIKELIHOOD OF INTERPERSONAL VIOLENCE

Three broad approaches to risk assessment have been discussed in the literature, clinical diagnoses of "dangerousness," psychological tests, and actuarial approaches.

Clinical Appraisals

The earliest empirical studies of risk assessment investigated the accuracy of clinical diagnoses of "dangerousness" among long-term male psychiatric patients who had been released into the community. Follow-ups based on official records such as arrest reports indicated that these diagnoses were accurate in only 20% (Steadman & Cocozza, 1974) to 35% (Kozol, Boucher, & Garafalo, 1972) of the cases. Recent studies using more sophisticated assessments and improved research methods now suggest that the accuracy of short-term clinical predictions of violence among acute psychiatric patients may be as high as 50% (Borum, 1996; Monahan, 1996; Otto, 1992). Hit rates of this magnitude can be useful in generalized risk assessment, but they are not very helpful in assessing an individual.

Psychological Tests

As a clinical psychologist for the Alameda County (California) Probation Department, the writer conducted a series of studies to determine which psychological tests or scales would best identify applicants for probation who were likely to assault citizens if released. It seemed then that it would be easy to compare the scores of offenders who had been convicted of extreme and moderately assaultive offenses with those of nonviolent criminals and normals. Once it was established which measure best *postdicted* criminal violence, it could be administered prospectively to determine its *predictive* validity. In the course of this research the MMPI, the California Psychological Inventory (CPI), the Rorschach, the Holtzman Inkblot Technique (HIT), the Thematic Apperception Test (TAT), and the Rosenzweig Picture-Frustration Study (P-F) were all studied using samples of assaultive and nonassaultive juvenile or adult offenders. None, not even an MMPI scale we ourselves derived, was able to differentiate the assaultive from the nonviolent offenders (Megargee, 1964, 1965; Megargee & Cook, 1967; Megargee & Mendelsohn, 1962).[3] The conclusion after reviewing the results of these and other studies in the literature was:

> Thus far no structured or projective test scale has been derived which, when used alone, will predict violence in the individual case in a satisfactory manner. Indeed, none has been developed which will adequately *post*dict, let alone *pre*dict, violent behavior. (Megargee, 1970, p. 145)

Otto (in press) recently wrote, ". . . This conclusion appears sound today."

This does not mean that these tests are useless; as we shall see, they can be very helpful in assessing some constructs relevant to aggression, such as hostility and inhibitions. But no test, *by itself*, can adequately differentiate aggressive from nonaggressive people. Aggression is simply too complex and multiply-determined a phenomenon to be assessed adequately by any single test.

Recall the Texas Tower massacre. Not just one but three different individuals committed homicide in Austin that day. One was the sniper, Charles Whitman, a highly controlled person suffering from an undiagnosed brain tumor who had murdered his mother and his wife the night before (Johnson, 1972). The other two were an armed civilian and a police officer who managed to gain access to the tower and kill Whitman. How could a single measure have accurately identified three such diverse men with such differing motivations?

Actuarial Devices

Empirical studies of factors associated with institutional adjustment and violence have produced a number of objective risk assessment devices designed to assist staff in making management, treatment, and release decisions. In prison settings, objective inmate classification tools have been devised by the Federal Bureau of Prisons, the California and Wisconsin Departments of Corrections, and the National Institute of Corrections (Austin, 1993; Brennan, 1987, 1993; Glaser, 1987), while the Province of Ontario uses the Level of Supervision Inventory (Bonta & Motiuk, 1985) to classify probationers and halfway-house residents. Although specific models differ, in general the more serious and extensive the criminal and correctional history, the greater the risk the offender is assessed as posing. An important benefit of these instruments is that they indicate that many low-risk prisoners could safely be assigned to less secure (and less expensive) correctional programs, whereas the subjective appraisals previously used had tended to overclassify many inmates (Solomon & Camp, 1993).

The devices developed in mental health settings are more specifically focused on predicting violence. In Canada, a series of instruments by Webster and his associates include the Dangerous Behavior Rating Scale based on the writer's "algebra of aggression model," (Webster & Menzies, 1993); the Violence Risk Assessment Guide (Harris, Rice, & Quinsey, 1993), a record-based assessment tool, which, when combined with a 10-item clinical scale (ASSESS-LIST; Webster & Polvi, 1995), comprises the Violence Prediction Scheme (Webster, Harris, Rice, Cormier, & Quinsey, 1994); the Spousal Assault Risk Assessment Guide (Kropp, Hart, Webster, & Eaves, 1994); and, most recently, the HCR-20 (Webster, Douglas, Eaves, & Hart, 1997). In the United States, risk assessment tools have been devised by Klassen and O'Connor (1988) and by the MacArthur Risk Assessment Study (Steadman et al., 1994).

These models consider such empirically identified risk factors as age; gender; socioeconomic status; the extensiveness and the nature of prior juvenile and adult criminal records, especially with respect to violent offenses; any history of experiencing or witnessing domestic violence; and the presence of certain specific psychiatric problems such as substance abuse, psychopathy, and paranoid ideation, as well as noncompliance with medication orders or treatment programs (Borum, 1996; Otto, 2000). These generalized risk assessment tools will be most useful in populations with high base rates for violent behavior and among clients with histories of previous violence. Research is needed to determine the degree to which predictors identified in one setting can validly be general-

ized to other settings; for example, whereas women in the general population are much less likely to be violent than men, female psychiatric inpatients are more apt to be violent than their male counterparts (Monahan, 1996; Otto, 2000).

Sources of Risk Assessment Data

Case history data are usually more informative than psychological test scores in risk assessment. Because aggressive behavior is external and occurs in an interpersonal context, it is noticed and documented more adequately than internal symptomatic behaviors. If the aggression results in criminal charges or a civil suit there may be extensive records available that can help the clinician determine where, when, and in what circumstances a person is or is not likely to commit certain types of aggressive acts directed at specific targets. D. M. Buss and Craik's (1983) "act frequency" approach to assessment can help the assessor delineate templates of aggressive or nonaggressive behaving in response to environmental contingencies. By comparing a client's current situation with these templates, the clinician can gain some insight into the likelihood of future recurrence of these behaviors.

Clinical interviews are useful not only for the content they elicit but also for the opportunity they provide to observe clients' emotional reactions. Are they quick-tempered? Do they have a sense of humor that helps them dissipate their anger?

Since aggressive behavior is interpersonal, friends, relatives, and co-workers who have observed a client's behavior in a variety of situations can provide a great deal of useful information. In institutional settings, psychologists should cultivate the people who actually spend a lot of time with the person being assessed. Nurses, aides, and fellow patients in mental health settings, and the correctional officers, teachers, work supervisors, and fellow inmates in correctional settings often know the clients' aggressive potential better than a therapist who sees them once a week. If people are feared by their peers, it should be taken seriously. In addition, medical charts showing the response to the administration, and, especially, the withdrawal, or medication and providing information on alcohol or substance abuse should be obtained.

In outpatient or community settings, an interview in the client's residence can be enormously informative. There is no substitute for experiencing the sights, sounds, and smells of the person's home environment. Alcohol and substance abuse by the client or other family members are more evident in a dwelling where you may see or smell the residue. In a home visit you are also better able to ascertain if the person has access to weapons and is skilled in their use. Be alert for gun cabinets, hunting photographs, trophies, and shooting magazines, and inquire about recreational activities.

THE CONCEPTUAL FRAMEWORK

When all the data have been collected—the records, reports, test protocols, observations, and so on—the conceptual framework makes its contribution. By specifying the complex determinants of aggressive behavior, it helps clinicians sort out and interpret the meaning of their observations. A brief overview of this "algebra of aggression" will be presented, after which each of its components will be discussed.

The Algebra of Aggression

Because most human behavior, including aggression, is performed on a fairly routine basis, it is easy to lose sight of its complex determinants. However, each individual act is the result of the interaction of many factors and dozens of unconscious choices. In the "algebra of aggression," we stop the action for a moment and analyze the determinants of a single act occurring at a specific point in time and in a particular situation.

At any given time, a person is confronted with dozens of implicit or explicit choices about how to behave; some acts may be aggressive, others nonaggressive. The aggressive responses

may be verbal or physical, mild or extreme, direct or indirect. How is this choice made?

According to the "algebra of aggression":

> Typically, a person selects the response that appears to offer the maximum satisfactions and the minimum dissatisfactions in that particular situation.
>
> This simple statement conceals a rapid but extremely complex internal bargaining process in which the capacity of a given response to fulfill many different drives and motives is weighed against the pain that might result from that response, as well as from the postponement of the satisfaction of other competing drives. . . . By means of this "internal algebra," the net strength of each possible response is calculated and compared with all other responses, and the strongest one is selected.
>
> What determines the net strength of a potential response? In the case of an aggressive or violent response, we can isolate several broad factors that interact to determine response strength. The first of these is *instigation to aggression*. Instigation to aggression is the sum of all the forces that motivate an individual to commit a violent or aggressive act. It includes both *intrinsic ("angry") instigation*, which is the conscious or unconscious wish to harm the victim in some fashion, and *extrinsic ("instrumental") instigation*, which is the yearning for other desirable outcomes which the aggressive act in question might accomplish, such as economic gain in the case of a robbery or perceived political benefits from an act of terrorism.
>
> The second major variable is *habit strength*, the extent to which the response has been rewarded or punished in the past. Other things being equal (which they rarely are), the more often a given aggressive act directed at a particular target has been successful in the past, or the more one has observed people aggressing successfully, the more likely one is to aggress in the future. Habit strength is especially relevant in the case of extrinsic or instrumental aggression.
>
> Instigation to aggression and habit strength both motivate people toward aggression. What stops them? Opposing the motivational factors is the third set of variables, namely *inhibitions against aggression*. They include all the reasons a person would refrain from a particular aggressive act directed at a particular target. Included are both moral prohibitions, which classify the particular act as wrong, and practical considerations, such as fear of retaliation or the possibility the act may fail in its objective. Inhibitions can be general or specific and can vary as a function of the act, the target, and the circumstances.
>
> Instigation, habit strength, and inhibitions are all personal characteristics, but behavior results from individuals interacting with their milieus. The fourth class of variables, *situational factors*, encompasses those external factors that may facilitate or impede aggressive behavior. These include *environments*, *settings*, *situations*, and *stimuli*. Among the external factors that might facilitate aggressive behavior are living under apartheid (environment), being in a war zone (setting), being present when a fight breaks out (situation), or a provocative gesture on the part of an antagonist (stimulus). Factors that might inhibit violence include living in a cloistered convent (environment), attending a symphony concert (setting), the presence of a police officer (situation), or having an opponent drop to his knees and raise his hands in supplication (stimulus). The common denominator of these events, according to Monahan and Klassen (1982), is that they all occur outside the skin.
>
> *Reaction potential*, the fifth and last major construct, consists of the net strength of a given response after the inhibitory factors have been balanced against the excitatory ones. A response will be blocked and cannot occur whenever the inhibitions exceed the instigation. A response is possible (i.e., has a positive reaction potential) if the forces favoring the aggressive response exceed those opposing it. However, all the possible responses must first compete with one another; the one with the highest reaction potential—that is, the greatest capacity to satisfy the most needs at the least cost—should be chosen. (Megargee, 1993, pp. 620–621)

Symbolically then, if

- $R_{j \cdot k}$ = reaction potential for a particular act j directed at target k
- $M_{j \cdot k}$ = motivation for a particular act j directed at target k
- H = habit strength
- $I_{j \cdot k}$ = inhibitions against act j directed at target k
- S_f = situational factors favoring or conducive to aggressing
- S_i = situational factors impeding or interfering with aggression

then the reaction potential for act "j" directed at target k is:

$$R_{j \cdot k} = [M_{j \cdot k} + H + S_f] - [I_{j \cdot k} + S_i]$$

Of course other acts directed at the same or at other targets will each have their own reaction potentials. For some, the inhibitory factors $[I_{j \cdot k} + S_i]$ will be stronger than the excitatory factors $[M_{j \cdot k} + H + S_f]$. When this occurs, the reaction potential for those responses is zero, so they are blocked and cannot occur. Perhaps the person was not angry enough, or the act in question might have been too extreme or inappropriate for that target under those circumstances. However, there may be some responses for which the excitatory factors $[M_{j \cdot k} + H + S_f]$ will exceed the inhibitions $[I_{j \cdot k} + S_i]$. All of these responses are possible; they will compete with one another and with other nonaggressive responses that may also be possible.

Given this brief overview of the conceptual framework, let us examine how it can be used in assessment and treatment.

ASSESSING THE PERSONAL FACTORS

The conceptual framework gives equal weight to personal and situational factors. The personal factors are instigation to aggression, habit strength, and inhibitions against aggression.

Intrinsic Instigation to Aggression

Depending on duration and intensity, intrinsic motives are referred to as *anger*, *hostility*, *rage*, or *hatred*. Obviously, the more intense and long-lasting the instigation, the easier it is to assess. Because anger and rage are transitory, they may not be observed in a particular session, so it is always advisable to see the client on two or more different occasions.

The primary questions clinicians must address are the amount and the direction of instigation to aggression. The secondary questions are the causes of this instigation and how the subject dissipates it. Or, more prosaically, "Is the client angry? At whom? Why? What does the client do about it?"

One of the best ways to assess anger and hostility is also the simplest: asking subjects, using a variety of synonyms, what makes them angry or annoyed, who aggravates them, what they would like to do when they are provoked, to whom they would like to do it, and what it is they actually do and why. Anger is a strong emotion; given the opportunity, many people will express it quite openly, even when it is not in their best interests to do so. Because anger is difficult to conceal, their friends and associates should also be able to comment on a client's temper.

Many self-report tests and scales have been constructed to assess anger and hostility. One of the most widely used, at least in research, is the Buss-Durkee Hostility Guilt Inventory (A. H. Buss & Durkee, 1957), which features seven scales designed to assess various aspects of hostility and aggression, plus one for assessing guilt over its expression, a potential measure of inhibitions. Since anger and hostility have been implicated as factors in cardiovascular and gastrointestinal disorders, a number of measures have been devised for use in behavioral medicine (Chesney & Rosenman, 1985). Spielberger (1996) has devised a State Trait Anger Expression Inventory (STAXI) with three standard scales, "Anger out" (Ax/out), "Anger in" (Ax/in), and "Anger control" (Ax/con), as well as two experimental subscales "Anger control in" (Ax/con-in), and "Anger control out" (Ax/con-out). Novaco's (1985) Anger scale focuses on the range of stimulus situations eliciting angry responses. MMPI-2 has an Anger content scale (Butcher, Graham, Williams, & Ben-Porath, 1990). Most of these scales are quite obvious and easily dissembled by clients motivated to conceal hostility. This is also true of the Rosenzweig Picture-Frustration Study, on which subjects indicate the responses cartoon figures would make to various frustrations (Rosenzweig, 1945; Rosenzweig, Fleming, & Clarke, 1947).

Less direct, and probably less valid, are those wide-band projective tests that include signs or indices purporting to reflect anger or hostility. These include symbolic content, such as anatomical responses or predatory animals, on the Rorschach (Phillips & Smith, 1953), and themes of fighting or conflict on apperceptive tests such as the TAT (Murstein, 1963). Situational factors such as the nature, tone, and purpose of the administration as well as the specific cards used strongly influence the responses obtained (Murstein, 1963). The less the TAT card suggests aggressive themes, the more weight one should give to their occurrence. Because the evidence for the validity of any single specific sign is weak (Megargee & Cook, 1967), clinicians assessing instigation to aggression should use projectives primarily to generate hypotheses to be tested with other data.

For both retrospective and prospective assessments it is important to determine the sources of clients' anger and hostility. Physiological sources include: (1) genetic factors that make some people more irritable than others; (2) certain diseases or conditions of the central nervous or endocrinological systems; (3) hormonal influences such as testosterone; (4) physical illness; (5) toxic factors and drugs; (6) fatigue; and (7) generalized arousal. Psychological sources include: (1) frustration; (2) physical or verbal attacks; and (3) territorial intrusions (Megargee, 1993). Motivation from all these different sources can add together.

The diagnosis of the source of the instigation can have direct implications for management and treatment. One of the best-established relationships in psychology is that frustration causes instigation to aggression (Dollard, Doob, Miller, Mowrer, & Sears, 1939). Does a client's frustration, and hence anger, stem from unreasonable expectations? If so, cognitive therapy designed to correct these beliefs might be effective. On the other hand, if the grievance is genuine and the anger justified, a better solution might be to help the client find a more constructive way of coping with the situation, such as seeking redress in the courts.

How people deal with anger or hostility is also important to assess. According to "catharsis" theory, intrinsic instigation should be depleted by attacking the source of the provocation. If this is impossible, angry individuals may reduce instigation by attacking someone else (displacement), choosing an alternative response (response substitution), or observing someone else's aggressive behavior, either in real life or the media (vicarious aggression). The overcontrolled-assaultive type lacks such mechanisms so that instigation accumulates to the point that violence erupts. Assertion training could be helpful for such people (Megargee, 1966, 1993).

Extrinsic (Instrumental) Instigation to Aggression

People aggress not only because they are angry, but also because aggression may help them attain other goals. As Al Capone reportedly said, "You can get much farther with a kind word and a gun than you can with a kind word alone" (Peter, 1977, p. 141). There is a vast array of possible extrinsic motives, both primary and secondary. They include personal gains and satisfactions, such as acquisition or enhancement of self-esteem; removal of problems, such as enemies or witnesses to crimes; and achieving social goals, such as power, or the respect of others; as well as the attainment of political or religious goals, as exemplified by the conflicts in the Middle East and elsewhere (Megargee, 1993, p. 530).[4] Clinicians should be alert to secondary as well as primary motivation. A reformed bank robber reported that although money had been the primary motive for his robberies, he had also enjoyed the excitement he experienced playing "cops and robbers" with real guns and cars.

Although psychologists working with individuals tend to focus on intrinsic motives such as anger and hostility, extrinsic motivation probably leads to as much if not more aggression, especially when we consider large-scale conflicts between groups and nations. Moreover, aggression is part of the job description for many people, including professional athletes, police officers, and military personnel.

Not being a personality trait, extrinsic motivation is much more difficult to assess than intrinsic. Lacking tests to assess instrumental motivation, clinicians must often rely on self-reports elicited in response to probing questions in an interview or infer how aggressive behav-

ior is reinforced from case history data. Obviously this works better in retrospective than prospective assessments.

If the aggressive behavior is truly instrumental, the intervention most likely to work is to show clients a more effective nonaggressive strategy to attain their goals. While this would not eliminate the aggressive response from their repertoires, it should make the nonaggressive alternative more likely to be chosen in the response competition.

Habit Strength

Reinforcement of aggressive responses increases habit strength, the second major personal factor leading to aggression.[5] Angry instigation is reinforced by the pain or discomfort inflicted on the victim; instrumental by attaining extrinsic goals by means of aggressive behavior. Reinforcement is strongest for the actual response and generalizes to others that are similar to it. The more a culture, society, or reference group approves of aggressive behavior, the more secondary reinforcement a person will receive.

Although habit strength is most readily acquired directly by reinforcement of an individual's own aggressive responses, people can also acquire habit strength indirectly by observing role models successfully engage in aggressive behavior. This may account in part for familial transmission of aggression, on the one hand, and how observing enactments of violence on television and other media can foster aggressive behavior, on the other (Bandura, 1973; Huesmann & Eron, 1986).

Extinction, that is continued performance of unreinforced aggressive responses, is the only way to eliminate habit strength once it is acquired. This is virtually impossible given the rewards for aggression in our culture, especially for boys. Although punishment may temporarily suppress aggressive behavior, it does not eliminate such responses from one's repertoire. Moreover, it may increase instigation. This is one reason prisons are often ineffective.

Of all the constructs in the "Algebra of Aggression," habit strength should be the easiest to assess, based as it is on the individual's reinforcement history as documented in the case history. One can infer strong habit strength from a long history of aggressive behavior, especially if it appears to have been successful. Habit strength is the variable that most strongly predicts aggression; the longer and stronger the history of aggression, the more likely an individual will behave similarly in the future. A history of aggressive behavior is a key element in diagnosing Megargee's (1966) undercontrolled assaultive type, and is central to the actuarial tools used in mental health and correctional risk assessment (Borum, 1996, Otto, 2000).[6] However, as in the case of the overcontrolled assaultive type, the fact that some people have no known history of aggression or violence is no guarantee that they will never aggress in the future.

Inhibitions Against Aggression

Opposing instigation and habit strength are internal inhibitions, which can be general or specific and can vary as a function of the aggressive act, the target, and the circumstances. These include pragmatic concerns, such as fears that an attack will fail or that retribution will be forthcoming, as well as moral prohibitions (i.e. taboos, conscience, superego) that classify this act directed at this target at this time under these circumstances as "wrong."

Inhibitions are by far the most difficult of the personal constructs to assess (Megargee, 1997). The mere absence of a history of aggressive behavior or failure to express animosity toward anyone does not necessarily indicate that people have strong inhibitions. A more parsimonious explanation would be that they lacked instigation.

To assess inhibitions directly, one must look for a circumstance in which a person is motivated to aggress and fails to do so. While such situations can be created in the laboratory, they are difficult to contrive in the clinic. Role playing, group therapy, and direct confrontations in an interview may afford opportunities to assess inhibitions directly. More often, clinicians have to ask people about what aggressive behaviors they find acceptable or unacceptable under various circumstances. Inhibitions are easier than anger to dissimulate, and potentially self-serving reports need to be cross-checked with the case history and reports by people who have had an opportunity to observe a client's behavior in provocative circumstances.

Often, however, clinicians must rely on indirect evidence such as personality test data. It is safest to use several possible sources and try to "triangulate" one's appraisal of inhibitions against aggression. Although Hartshorne and May's (1928, 1929) classic studies demonstrated the specificity of moral prohibitions, other things being equal, well-socialized people with strong values are more likely to have inhibitions against those forms of aggression that are censured by their culture than those who are poorly socialized. Among the best measures of socialization and values available are the California Psychological Inventory's (CPI) Socialization (So), Responsibility (Re) and Self Control (Sc) scales (Gough & Bradley, 1996). These scales, which have been validated in many cultures, assess the degree to which individuals have assimilated (So) and understand (Re) their cultural values, and their ability to control their behavior and abide by these precepts (Sc) (Megargee, 1972).

Although the writer prefers using the CPI among "normal" populations, with more deviant or pathological groups MMPI-2 is preferable (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). A clinically elevated Overcontrolled Hostility (*O-H*) scale score suggests a conflict between expression or suppression of hostility, but low scores are meaningless (Megargee et al., 1967). Elevations and high points on basic scales F, 4, 6, 8, and 9; on supplementary scale MAC-R; and on content scale ASP suggest *deficient* inhibitions and controls (Megargee, Mercer, & Carbonell, 1999). Obviously, the more of these scales that are elevated and, assuming the profile is valid, the higher the elevations, the stronger the evidence.

No one is more lacking in inhibitions than the psychopath, and Hare's (1991) Psychopathy Checklist-Revised (PCL-R) has been related to violent criminal behavior in a number of studies. Reviewing the literature on the PCL-R and risk assessment, Hart (1996, p. 64) concluded, ". . . predictions of violence using the PCL-R are considerably better than chance, albeit far from perfect."

Projective tests can also be used to assess deficient controls and poor socialization. Consider a Rorschach or TAT administration as a structured interpersonal interaction between the client and the clinician for a mutually agreed-on purpose. In most instances in which aggression is being assessed, it is in the subject's best interest to cooperate, take the examination seriously, and produce a protocol that does not violate the implicit social norms governing the situation. One can therefore hypothesize that subjects who produce inappropriate or deviant responses may be lacking suitable inhibitions or controls. On the Rorschach, for example, uncontrolled color responses that disregard the form or shape of the blot, and blatant sexual, anatomical, gruesome, or gory content suggest that the respondent doesn't understand or disregards social conventions. On the other hand, a record with many well-differentiated responses with good form, using frequently interpreted blot areas and including a good number of Popular (P) responses indicates the respondent understands the proprieties and is willing and able to abide by them, at least in the context of the inkblot examination.

Compared with the Rorschach, there is much greater interexaminer variability in the administration, scoring, and interpretation of the TAT and other apperceptive measures, making generalizations more difficult. If respondents *fail* to give aggressive themes in response to cards or scenes that strongly suggest aggression or conflict, then one might infer inhibitions against the expression of aggression, at least in the context of the examination (Megargee, 1967).

Another strategy is to examine the case history to determine whether the person came from a home conducive to strong moral development. The conditions associated with socialization are well known, even if the actual mech-

anisms by which values are introjected may be poorly understood:

> A warm nurturing environment in which the children form a close bond of affection and respect with their caregivers combined with fair, consistent discipline is most conducive to developing internal controls. The more the agents of socialization, first the parents, and later the neighborhood, the church, and the school, share and enforce a strong common set of values, and the more they live up to and consistently exemplify these principles in their everyday lives, the more likely it is that the children will incorporate them. . . . On the other hand . . . parental absence and disharmony, inconsistency, rejection, abuse and poor role models are associated with problems in developing values and controlling behavior. . . . (Megargee 1997, p. 596)

Although there are obviously many exceptions, the more the case history indicates that the client grew up in a milieu conducive to socialization, the more reasonable it is to hypothesize a normal set of values.

In addition to values stipulating that certain types of aggression are wrong, inhibitions are also fostered by empathy or compassion for the potential victim. The CPI Empathy (Em) scale, as well as its Sociability (Sy) and Social Presence (Sp) scales might be used to estimate this construct. Similarly, TAT themes emphasizing affiliation and warm, mutually satisfying interpersonal relations, along with an emphasis on Human Movement and controlled color on the Rorschach, would all suggest a capacity for relating to others that could strengthen inhibitions against aggression.

Unfortunately, it is much easier to decrease inhibitions than it is to foster them. Physiological factors that reduce inhibitions included: (1) injuries or diseases affecting the central nervous system; (2) certain endocrinological disorders; and (3) the chemical actions of disinhibiting substances such as alcohol. Psychological causes of inadequate inhibitions include: (1) failure to develop adequate inhibitions because of deficient socialiation or abuse during childhood; and (2) being exposed to role models who approve of violence. Inhibitions against violence can also be temporarily overcome by conflicting values (i.e., it is wrong to attack someone but right to defend your honor), rationalizations, and peer pressure. An adequate assessment of inhibitions must explore whether people's inhibitions against aggression are easily overcome.

SITUATIONAL FACTORS AND EXTERNAL CONDITIONS

How Situational Factors Influence Aggression

Aggression results from the interaction of the personal factors discussed above with external conditions that can either facilitate or impede aggressive behavior or violence. These situational factors can be divided into *environments, settings, situations,* and *stimuli* to demarcate a rough continuum from widespread to focal influences. The boundaries between these terms are not significant; the point is that a wide range of external factors and events helps determine whether or not an aggressive act takes place. Situational influences involve times as well as places; much more violence occurs in New York City's Central Park between 2:00 and 4:00 A.M. on Sunday morning than from 2:00 to 4:00 P.M. on Sunday afternoon. Of course, personality and situational factors are not independent. Those peaceful people who flock to Central Park on a sunny afternoon generally shun it in the early morning hours, leaving it to the predators and police.

The circumstances and stimuli to which clients are exposed and the conditions in which they live are extremely important determinants of aggression. Cognitive expectancies interacting with situational realities can cause frustration, the major source of intrinsic instigation (Dollard et al., 1939). Environmental factors can moderate the relation of personality factors, such as empathy, to aggression (Rose & Feshbach, 1990). Given strong enough threat or provocation, almost anyone may become violent.

Since this book is about personality assessment, space does not permit a discussion of all the external factors that influence aggression.

Among those that have been emphasized in the literature are ambient temperature (Baron & Ransberger, 1978; Megargee, 1977); architectural design (Newman, 1972); crowding (Megargee, 1977; Russell, 1983); the family, peer, and job environments; the availability of alcohol and of potential victims (Monahan & Klassen, 1982); and the behavior of antagonists, victims, associates, and bystandards (Megargee, 1993). Obviously the political climate is a major factor, as people from Baghdad, Beijing, Beirut, Bethlehem, and Bosnia will attest. Probably the most widely discussed variable is the availability of weapons by either or both antagonists (Cook, 1982; MacDonald, 1975; Monahan & Klassen, 1982).

Assessing Situational Factors

In both prospective and retrospective assessments, clinicians must consider situational as well as personality factors Bjørkly (1993, 1994) devised a situational "Scale for the Prediction of Aggression and Dangerousness in Psychotic Patients," which is designed to predict the frequency and intensity of psychiatric patients' aggressive behavior in 29 different situations, both on the ward and off. Recent research (Bjørkly, Havik, & Løberg, 1996) attested to its interrater reliability, but its validity has yet to be established, as does its applicability outside of Norway. Monahan and Klassen (1982) suggested that we investigate the circumstances that led people to commit aggressive acts in the past and determine whether they correspond to the situations they are likely to encounter in the future. We might also ascertain the environmental and situational factors associated with peaceful periods to help guide recommendations for future treatment or environmental manipulation.

The less we know about the settings in which clients will find themselves or the situations and stimuli they will face, the more difficult it is to make accurate predictions. Only in institutional settings, such as prisons and inpatient mental health units, where the environmental factors are relatively constant for all subjects, does it appear possible to make reasonably accurate predictions based only on personality factors (McGuire & Megargee, 1976).

RESPONSE COMPETITION AND REACTION POTENTIAL

Any violent act will be blocked if inhibitory factors exceed motivating ones. However, if all the forces favoring a particular violent response exceed those opposing it, that act is *possible,* at least as long as those conditions prevail. Before it can be carried out, all the possible responses, both violent and nonviolent, must compete with one another; the one with the highest *reaction potential*—that is, the greatest capacity to satisfy the most needs at the least cost—should be chosen.

Assessing Reaction Potential

In a retrospective analysis of aggressive behavior, it is obvious that the act that was actually performed had the highest reaction potential. If that act was socially or personally undesirable, and the clinical goal is to decrease the likelihood of a recurrence, other, more satisfactory alternatives should be investigated. Why were these acts not selected? Were the inhibitions too great? Their chance of success too low? Did cognitive beliefs suggest that the aggressive act was the more honorable alternative? Did friends or onlookers encourage the deviant behavior? The answers to such questions might indicate how the reaction potential of the objectionable behavior may be decreased and that of more acceptable responses increased.

In prospective assessments, the clinician's task is to try to determine the range of possible responses and estimate their relative reaction potential. Situational circumstances, such as whether or not certain interventions are attempted, may be critical. Instead of affixing a blanket label, such as "dangerous," these analyses should help clinicians make more sophisticated contingent predictions of the likelihood that clients will engage in specified aggressive behaviors under different sets of circumstances.

HIT RATES AND BASE RATES

The basic thesis of this chapter is that aggression is a complex, multiply-determined, interpersonal interaction involving both personal and situational factors. To assess the risk of aggressive behavior, one must appreciate and consider its intricate determinants.

Because these elements are difficult to assess, it is inevitable that errors will occur. Although investigating the accuracy of predictions of aggressive behavior has inherent difficulties, the best recent studies seem to indicate that it is possible to achieve a hit rate of one in two; that is, for every two people clinicians identify as having a high likelihood of engaging in the type of aggressive behavior being studied, one will actually do so (Borum, 1996; Otto, 1992). This hit rate is quite adequate for some purposes, but not for others (Shah, 1977).

The effect of such errors can be magnified by the low base rates for the occurrence of aggressive behavior in many settings. The closer the base rate is to 50%, the higher the possible hit rate (Finn & Kamphuis, 1995; Meehl & Rosen, 1955). As the frequency of the behavior decreases, the effects of even moderate false-positive rates are greatly increased. Therefore, in addition to everything else, clinicians making prospective assessments must be aware of the base rates for aggressive behavior in their settings and the impact of these rates on the accuracy of their predictions (Meehl & Rosen, 1955; Megargee, 1976; Monahan, 1981).

NOTES

1. These are the possible consequences of being labeled "dangerous" in societies governed by laws and due process. Elsewhere being labeled "dangerous" can lead to simpler, more direct interventions, such as execution. In this century, the regimes of Joseph Stalin, Adolf Hitler, Idi Amin, and Pol Pot, to name a few, serve as examples.

2. For example, few psychologists are aware of the fact that threatening the life of the President of the United States is itself a crime (18 USC 871). Therefore, psychologists who fail to report such threats could themselves be accused of misprision of a felony.

3. Some scales were able to discriminate the criminals from the noncriminals, but this was of little use in an agency dealing exclusively with criminal offenders.

4. See Megargee (1982) for a detailed discussion of extrinsic motives.

5. Whereas the construct of intrinsic instigation originated in frustration/aggression theory (Dollard et al., 1939) and extrinsic instigation was suggested by A. H. Buss's (1961) writings, the concept of habit strength comes from Hull's (1952) model of motivation. Habit strength is the major reason the algebra of aggression postulates an additive rather than a multiplicative model. In theory, before a person has ever been reinforced for aggressing, his or her habit strength should be zero. In a multiplicative model, this would make it theoretically impossible for anyone to aggress for the first time.

6. Of course, such a history also suggests low inhibitions and high instigation.

REFERENCES

American Association for Correctional Psychology, Standards Committee (2000). Standards for psychological services in jails, prisons, correctional facilities, and agencies (2nd ed.). *Criminal Justice and Behavior, 27,* 433–494.

American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist, 47,* 1597–1628.

Austin, J. (1993). Objective prison classification systems: A review. In American Correctional Association (Ed.), *Classification: A tool for managing today's offenders* (pp. 108–123). Laurel, MD: American Correctional Association.

Bandura, A. (1973). *Aggression: A social learning analysis.* Englewood Cliffs, NJ: Prentice-Hall.

Baron, R. A. (1977). *Human aggression.* New York: Plenum.

Baron, R., & Ransberger, V. (1978). Ambient temperature and the occurrence of collective violence: The "long hot summer" revisited. *Journal of Personality and Social Psychology, 36,* 351–360.

Bennett, B. E., Bryant, B. K., Vandenbos, G. R., & Greenwood, A. (1990). *Professional liability and risk management.* Washington, DC: American Psychological Association.

Bjørkly, S. (1993). The Scale for the Prediction of Dangerousness and Aggression in Psychotic Patients: An introduction. *Psychological Reports, 73,* 1363–1377.

Bjørkly, S. (1994). The Scale for the Prediction of Dangerousness and Aggression in Psychotic Patients: A prospective pilot study. *Criminal Justice and Behavior, 21,* 341–356.

Bjørkly, S., Havik, O. E., & Løberg, T. (1996). The interrater reliability of the Scale for the Prediction of Dangerousness and Aggression in Psychotic Patients (PAD). *Criminal Justice and Behavior, 23,* 440–454.

Bohn Jr., M. J. (1979). Inmate classification and the reduction of violence. In *Proceedings of the 109th Annual Congress of Correction* (pp. 63–69). College Park, MD: American Correctional Association.

Bonta, J, & Motiuk, L. L. (1985). Utilization of an interview-based classification instrument: A study of correctional halfway houses. *Criminal Justice and Behavior, 12,* 333–352.

Borum, R. (1996). Improving the clinical practice of violence risk assessment: Technology, guidelines, and training. *American Psychologist, 51,* 945–956.

Borum, R., Fein, R., Vossekuil, B., & Berglund, J. (1999). Threat assessment: Defining an approach for evaluating risk of targeted violence. *Behavioral Sciences and the Law, 17,* 323–337.

Brennan, T. (1987). Classification: An overview of selected methodological issues. In D. M. Gottfredson & M. Tonry (Eds.), *Prediction and classification in criminal justice decision making* (pp. 201–248). Chicago: University of Chicago Press.

Brennan, T. (1993). Risk assessment: An evaluation of statistical classification methods. In American Correctional Association (Ed.), *Classification: A tool for managing today's offenders* (pp. 46–70.) Laurel, MD: American Correctional Association.

Brodsky, S. L. (1991). *Testifying in court: Guidelines and maxims for the expert witness.* Washington, DC: American Psychological Association.

Buss, A. H. (1961). *The psychology of aggression.* New York: Wiley.

Buss, A. H., & Durkee, A. (1957). An inventory for assessing different kinds of hostility. *Journal of Consulting Psychology, 21,* 343–348.

Buss, D. M., & Craik, K. H. (1983). The act-frequency approach to personality. *Psychological Review, 96,* 105–126.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A. M., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring.* Minneapolis: University of Minnesota Press.

Butcher, J. N., Graham, J. R., Williams, C., & Ben-Porath, Y. S. (1990). *Development and use of the MMPI-2 content scales.* Minneapolis: University of Minnesota Press.

Chesney, M. A., & Rosenman, R. H. (Eds.). (1985). *Anger and hostility in cardiovascular and behavior disorders.* Washington, DC: Hemisphere.

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review, 44,* 588–608.

Cook, P. I. (1982). The role of firearms in violent crime: An interpretative review of the literature. In M. E. Wolfgang & N. A. Weiner (Eds.), *Criminal violence* (pp. 236–291). Beverly Hills, CA: Sage.

Cunningham, M. D., & Reidy, T. J. (1999). Don't confuse me with the facts: Common errors in risk assessment at capital sentencing. *Criminal Justice and Behavior, 26,* 20–43.

Dollard, J., Doob, L. W., Miller, N. E., Mowrer, O. H., & Sears, R. R. (1939). *Frustration and aggression.* New Haven, CN: Yale University Press.

Finn, S. E., & Kamphuis, J. H. (1995). What a clinician needs to know about base rates. In J. N. Butcher (Ed.), *Clinical personality assessment: Practical approaches* (pp. 224–235). New York: Oxford University Press.

Glaser, D. (1987). Classification for risk. In D. M. Gottfredson & M. Tonry (Eds.), *Prediction and classification in criminal justice decision making* (pp. 249–291). Chicago: University of Chicago Press.

Gough, H. G., & Bradley, P. (1996). *CPI manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.

Hare, R. D. (1991). *The Hare Psychopathy Checklist-Revised.* Toronto: Multi-Health Systems.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior, 20,* 315–335.

Hart, S. (1996). Psychopathy and risk assessment. In D. J. Cooke, A. E. Forth, J. Newman, & R. D. Hare (Eds.), *International perspectives on psychopathy, Issues in Criminological and Legal Psychology, No. 24* (pp. 63–67.) Leicester, England: The British Psychological Society, Division of Criminological and Legal Psychology.

Hartshorne, H., & May, M. A. (1928). *Studies in the nature of character. Vol. 1: Studies in deceit.* New York: Macmillan.

Hartshorne, H., & May, M. A. (1929). *Studies in the nature of character. Vol. 2: Studies in service and self control.* New York: Macmillan.

Heilbrun, K., & Heilbrun, A. B. (1995). Risk assessment with the MMPI-2 in forensic evaluations. In Y. S. Ben-Porath, J. R. Graham, G. C. N. Hall, R. D. Hirschman, & M. S. Zaragoza (Eds.),

Forensic applications of MMPI-2 (pp. 127–159). Thousand Oaks, CA: Sage.

Huesmann, L. R., & Eron, L. D. (1986). *Television and the aggressive child: A cross-national comparison*. Hillsdale, NJ: Erlbaum.

Hull, C. L. (1952). *A behavior system*. New Haven, CN: Yale University Press.

Johnson, R. N. (1972). *Aggression in man and animals*. Philadelphia: W.B. Saunders.

Klassen, D., & O'Connor, W. A. (1988). A prospective study of predictors of violence in adult male mental patients. *Law and Human Behavior, 12*, 143–158.

Kozol, H., Boucher, R., & Garafalo, R. (1972). The diagnosis and treatment of dangerousness. *Crime and Delinquency, 18*, 371–392.

Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1994). *Manual for the Spousal Assault Risk Assessment guide*. Vancouver: British Columbia Institute on Family Violence.

MacDonald, J. (1975). *Armed robbery: Offenders and their victims*. Springfield, IL: Charles C. Thomas.

McGuire, J. S., & Megargee, E. I. (1976, September). Prediction of dangerous and disruptive behavior in a federal institution for youthful offenders. In C. S. Moss (Chair), *Objective studies on violence in correctional settings*. Symposium presented at the meeting of the American Psychological Association, Washington, DC.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of certain psychometric signs, patterns, and cutting scores. *Psychological Bulletin, 52*, 194–216.

Megargee, E. I. (1964, April). *The utility of the Rosenzweig Picture-Frustration Study in detecting assaultiveness among juvenile delinquents*. Paper presented at the meeting of the Southwestern Psychological Association, San Antonio, TX.

Megargee, E. I. (1965). Relation between Barrier scores and aggressive behavior. *Journal of Abnormal and Social Psychology, 70*, 307–311.

Megargee, E. I. (1966). Undercontrolled and overcontrolled personality types in extreme antisocial aggression. *Psychological Monographs, 80* (3, whole No. 611).

Megargee, E. I. (1967). Hostility on the TAT as a function of defensive inhibition and stimulus situation. *Journal of Projective Techniques and Personality Assessment 31*(4), 73–79.

Megargee, E. I. (1970). The prediction of violence with psychological tests. In C. D. Spielberger (Ed.), *Current topics in clinical and community psychology* (Vol. 2, pp. 97–156). New York: Academic Press.

Megargee, E. I. (1972). *The California Psychological Inventory handbook*. San Francisco: Jossey-Bass.

Megargee, E. I. (1976). The prediction of dangerous behavior. *Criminal Justice and Behavior, 3*, 3–21.

Megargee, E. I. (1977). The association of population density, reduced space and uncomfortable temperatures with misconduct in a prison community. *American Journal of Community Psychology, 5*, 289–298.

Megargee, E. I. (1981). Methodological problems in the prediction of violence. In J. R. Hays, T. K. Roberts, & K. S. Solway (Eds.), *Violence and the violent individual* (pp. 179–191). New York: S. P. Scientific and Medical Books.

Megargee, E. I. (1982). Psychological correlates and determinants of criminal violence. In M. Wolfgang & N. Weiner (Eds.), *Criminal violence* (pp. 81–170). Beverly Hills, CA: Sage.

Megargee, E. I. (1993). Aggression and violence. In H. Adams and P. Sutker (Eds.), *Comprehensive handbook of psychopathology* (2nd. ed., pp. 617–644). New York: Plenum.

Megargee, E. I. (1997). Internal inhibitions and controls. In S. R. Briggs, R. Hogan, & W. H. Jones (Eds.), *Handbook of personality psychology* (pp. 581–614). New York: Academic Press.

Megargee, E. I., & Bohn Jr., M. J., with Meyer Jr., J., & Sink, F. (1979). *Classifying criminal offenders: A new system based on the MMPI*. Beverly Hills, CA: Sage.

Megargee, E. I., & Cook, P. E. (1967). The relation of TAT and inkblot aggressive content scales with each other and with criteria of overt aggressiveness in juvenile delinquents. *Journal of Projective Techniques and Personality Assessment, 31*(1), 48–60.

Megargee, E. I., Cook, P. E., & Mendelsohn, G. A. (1967). Development and validation of an MMPI scale of assaultiveness in overcontrolled individuals. *Journal of Abnormal Psychology, 72*, 519–528.

Megargee, E. I., & Mendelsohn, G. A. (1962). A cross-validation of 12 MMPI indices of hostility and control. *Journal of Abnormal and Social Psychology, 65*, 431–438.

Megargee, E. I., Mercer, S. J., & Carbonell, J. L. (1999). MMPI-2 with male and female state and federal prison inmates *Psychological Assessment, 11*, 177–185.

Meyers, C. J. (1986). The legal perils of psychotherapeutic practice: The farther reaches of the duty to warn. In L. Everstine & D. S. Everstine (Eds.), *Psychotherapy and the law*. New York: Grune & Stratton.

Monahan, J. (Ed.) (1972). *Who is the client? The ethics of psychological intervention in the criminal justice system*. Washington, DC: American Psychological Association.

Monahan, J. (1981). *The clinical prediction of violent behavior*. Rockville, MD: National Institute of Mental Health.

Monahan, J. (1996). Violence prediction: The past twenty years. *Criminal Justice and Behavior, 23*, 107–120.

Monahan, J., & Klassen, D. (1982). Situational approaches to understanding and predicting violent behavior. In M. E. Wolfgang & N. A. Weiner (Eds.), *Criminal violence* (pp. 292–319). Beverly Hills, CA: Sage.

Murstein, B. I. (1963). *Theory and research on projective techniques (emphasizing the TAT)*. New York: Wiley.

Newman, O. (1972). *Defensible space*. New York: Macmillan.

Novaco, R. W. (1985). Anger and its therapeutic regulation. In M. A. Chesney & R. H. Rosenman (Eds.), *Anger and hostility in cardiovascular and behavior disorders* (pp. 203–226). Washington, DC: Hemisphere.

Otto, R. (1992). The prediction of dangerous behavior: A review and analysis of "second generation" research. *Forensic Reports, 5*, 103–133.

Otto, R. (2000). Assessing and managing violence risk in outpatient settings. *Journal of Clinical Psychology, 56*, 1239–1262.

Peter, L. J. (1977). *Peter's quotations: Ideas for our time*. New York: William Morrow.

Phillips, L., & Smith, J. (1953). *Rorschach interpretation: Advanced technique*. New York: Grune and Stratton.

Pope, K. S., Butcher, J. N., & Seelen, J. (1993). *MMPI/ MMPI-2/ MMPI-A in court: Assessment, testimony and cross-examination for expert witnesses and attorneys*. Washington, DC: American Psychological Association.

Rose, A., & Feshbach, N. (1990, June). *Empathy and aggression revisited: The effects of context*. Paper presented at the IXth Biennial World Meeting of the International Society for Research on Aggression, Banff, Alberta.

Rosenzweig, S. (1945). The picture-association method and its application in a study of reaction to frustrations. *Journal of Personality, 14*, 3–23.

Rosenzweig, S., Fleming, E. F., & Clarke, H. J. (1947). Revised scoring manual for the Rosenzweig Picture-Frustration Study. *Journal of Psychology, 24*, 165–208.

Russell, G. W. (1983). Crowd size and density in relation to athletic aggression and performance. *Social Behavior and Personality, 11*, 9–15.

Shah, S. (1977). Dangerousness. Some definitional, conceptual and public policy issues. In B. Sales (Ed.), *Perspectives in law and psychology*. New York: Plenum.

Solomon, L., & Camp, A. T. (1993). The revolution in correctional classification. In American Correctional Association (Ed.), *Classification: A tool for managing today's offenders* (pp. 1–16). Laurel, MD: American Correctional Association.

Spielberger, C. D. (1996). *State-Trait Anger Expression Inventory: STAXI professional manual*. Odessa, FL.: Psychological Assessment Resources.

Steadman, H., & Cocozza, J. (1974). *Careers of the criminally insane*. Lexington, MA: Lexington.

Steadman, H., Monahan, J., Appelbaum, P., Grisso, T., Mulvey, E., Roth, L., Robbins, P., & Klassen, D. (1994). Designing a new generation of risk assessment research. In J. Monahan & H. Steadman (Eds.), *Violence and mental disorder: Developments in risk assessment* (pp. 297–318). Chicago: University of Chicago Press.

Stone, A. A. (1975). *Mental health and the law: A system in transition*. Washington, DC: U. S. Government Printing Office.

Stone, A. A. (1986). Vermont adopts *Tarasoff*: A real barn-burner. *American Journal of Psychiatry, 143*, 352–355.

Tarasoff v. Regents of the University of California, 131 Cal. Rptr. 14, 551 P.2d 334 (1976).

Toch, H. (1969). *Violent men*. Chicago: Aldine.

VandenCreek, L., & Knapp, S. (1989). *Tarasoff and beyond: Legal and clinical considerations in the treatment of life-endangering patients*. Sarasota, FL: Practitioner's Resource Series.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart., S. D. (1997). *HCR-20: Assessing risk for violence* (Version 2). Burnaby, BC: Mental Health, Law, & Policy Institute, Simon Frazier University.

Webster, C. D., Harris, G. T., Rice, M. E., Cormier, C., & Quinsey, V. L. (1994). *The Violence Prediction Scheme: Assessing dangerousness in high risk men*. Toronto: University of Toronto.

Webster, C. D., & Menzies, R. J. (1993). Supervision in the deinstitutionalized community. In S. Hodgins (Ed.), *Mental disorder and crime* (pp. 22–38). Newbury Park, CA: Sage.

Webster, C. D., & Polvi, N. H. (1995). Challenging assessments of dangerousness and risk. In J. Ziskind (Ed.), *Coping with psychiatric and psychological testimony* (pp. 221–240). Marina del Rey, CA: Law and Psychology Press.

4.03
Diagnostic Models and Systems

ROGER K. BLASHFIELD
Auburn University, AL, USA

4.03.1 INTRODUCTION 57
4.03.2 PURPOSES OF CLASSIFICATION 58
4.03.3 DEVELOPMENT OF CLASSIFICATION SYSTEMS IN THE USA 59
4.03.4 KRAEPLIN 62
4.03.5 EARLY *DSMS* AND *ICDS* 62
4.03.6 NEO-KRAEPELINIANS 63
4.03.7 *DSM-III*, *DSM-III-R*, AND *DSM-IV* 64
4.03.8 *ICD-9*, *ICD-9-CM*, AND *ICD-10* 67
4.03.9 CONTROVERSIES AND ISSUES 69
4.03.9.1 Organizational Models of Classification 69
4.03.9.2 Concept of Disease 72
4.03.9.3 Two Views of a Hierarchical System of Diagnostic Concepts 75
4.03.9.4 Problem of Diagnostic Overlap/Comorbidity 76
4.03.10 CONCLUDING COMMENTS 78
4.03.11 REFERENCES 78

4.03.1 INTRODUCTION

To all the People and Inhabitants of the United States and all the outlying Countries, Greetings:
I, John Michler, King of Tuskaroras, and of all the Islands of the Sea, and of the Mountains and Valleys and Deserts; Emperor of the Diamond Caverns, and Lord High General of the Armies thereof; First Archduke of the Beautiful Isles of the Emerald Sea; Lord High Priest of the Grand Lama, etc., etc., etc.: Do issue this my proclamation. Stand by and hear, for the Lord High Shepherd speaks. No sheep have I to lead me around, no man have I to till me the ground, but the sweet little cottage is all of my store, and my neat little cottage has ground for the floor. No children have I to play me around, no dog have I to bark me around, but the three-legged stool is the chief of my store, and my neat little cottage has ground for the floor.
Yea, verily, I am the Mighty King, Lord Archduke, Pope, and Grand Sanhedrim, John Michler. None with me compare, none fit to comb my hair, but the three-legged stool is the chief of my store, and my neat little cottage has ground for the floor. John Michler is my name. Selah!
I am the Great Hell-Bending Rip-Roaring Chief of the Aborigines! Hear me and obey! My breath overthrows mountains; my mighty arms crush the everlasting forests into kindling wood; I am the owner of the Ebony Plantations; I am the owner of all the mahogany groves and of all the satin-wood; I am the owner of all the granite; I am the owner of all the marble; I am the owner of all the owners of Everything. Hear me and obey! I, John Michler, stand forth in the presence of the Sun and of all the Lord Suns and Lord Planets of the Universe, and I say, Hear me and obey! I, John Michler, on this eighteenth day of August, 1881, do say, Hear me and obey! for with me none can equal, no, not one, for the three-legged stool is the chief of my store,

and my neat little cottage has ground for the floor. Hear me and obey! Hear me and obey! John Michler is my name.
John Michler, First Consul and Dictator of the World, Emperor, Pope, King and Lord High Admiral, Grand Liconthropon forever! (Hammond, 1883, pp. 603–604)

A physician in private practice in New York City reported that a man brought his brother, John Michler, to see him. John was acting strangely, and his brother wanted to know what to do. The brother gave the physician a proclamation that John Michler had written (see above). Clearly, to most observers, there would be no question that John Michler was "crazy." However, what is the diagnosis of John Michler?

When this proclamation was shown to mental health professionals, the most common diagnostic possibilities that are mentioned are schizophrenia and bipolar disorder (manic episode). What do these various diagnoses mean? Why did clinicians not assign a diagnosis of narcissistic personality disorder to this patient? Certainly this man would fit the vernacular meaning of self-centered and self-aggrandizing that is often associated with a narcissistic personality. How is a manic episode differentiated from schizophrenia? What does it mean to say that Michler appears to be psychotic? Does that diagnosis mean that he has some type of disease that has affected part of his brain, or does it suggest that his childhood was so unusual and abnormal that he has developed a strange and unusual way of coping with life?

4.03.2 PURPOSES OF CLASSIFICATION

The vernacular word that was used to describe John Michler was "crazy." This word is frequently used in descriptions of persons who have mental disorders. The reason for the applicability of this word is that one common feature of most psychiatric patients is that their behaviors are statistically abnormal. That is, psychiatric patients behave in ways that are deviant. Their interpersonal actions are not expected within the cultural context of our society.

Classification is a fundamental human activity that people use to understand their world. For instance, a classification of animals is helpful when understanding the variations among diverse forms of living organisms. In forming a classification, the classifier is attempting to use observed similarities and differences among the things being classified, so as to find some order among those things. In psychopathology, the general goal of classification is an attempt to use similarities and differences among people who behave in deviant and socially abnormal ways in order to understand their behaviors.

More specifically, there are five purposes to the classification of mental disorders: (i) forming a nomenclature so that mental health professionals have a common language; (ii) serving as a basis of information retrieval; (iii) providing a short-hand description of the clinical picture of the patient; (iv) stimulating useful predictions about what treatment approach will be best; and (v) serving as a concept formation system for a theory (or theories) of psychopathology.

The first reason to have a classification system, providing a nomenclature, is the most fundamental (World Health Organization, 1957). At a minimum, a classification system provides a set of nouns for clinicians to use to discuss their objects of interest—people. Thus, a nomenclature is a set of terms that refer to groups of people that mental health professionals see in their various professional roles.

The second reason, information retrieval, has a pragmatic focus on how well a classification organizes a research literature, so that clinicians and scientists can search for information that they need (Feinstein, 1967). In biology, there is a dictum: "the name of a plant is the key to its literature" (Sneath & Sokal, 1973). The same is true in the area of mental disorders. If a student clinician is assigned a patient who attempts to control weight by inducing vomiting, the name bulimia becomes important for helping the student locate the literature about this disorder in books, journal articles, and even on the internet.

The third reason for having a classification is description. There are many ways of creating classifications that could satisfy the first two purposes. For instance, clinicians could decide to classify all of their patients on eye color. Using eye color would allow clinicians to have a language to discuss patients ("I have seen 17 brown eyed, eight blue eyed, and four mixed eye color patients in the last month."). Also, eye color categories could be used as names to store information about patients. However, using eye color to classify patients would not be a satisfactory solution to either researchers or clinicians because patients with similar eye colors are unlikely to have similar symptom patterns. To meet the purpose of description, patients are grouped on the basis of similarity. Patients who have the same diagnosis are expected to be relatively similar in terms of symptoms (Lorr, 1966). In addition, these patients should be dissimilar when compared to patients with different diagnoses (Everitt, 1974). In the case of John Michler, if he is having a manic episode of a bipolar disorder, then we would expect that Michler's brother would report that John had been spending large sums of money that he did not have, that his speech was extremely rapid, and that his sleep pattern was markedly disturbed. In contrast, his brother is unlikely to report that Michler usually sat around the house with an emotionless, cold, detached interpersonal style and that he would tell his brother about voices in his head that were in discussion about Michler's behaviors. The latter symptoms typically occur in individuals with schizophrenia. Thus, diagnoses become descriptive short-hand names for clusters of co-occurring symptoms.

The fourth purpose is prediction. This purpose typically involves two types of inferences: (i) predicting the course of the patient's condition if there is no treatment or intervention ("Diagnosis is prognosis" as stated by Woodruff, Goodwin, & Guze, 1974); and (ii) predicting which treatment approach will be most effective with the patient (Roth & Fonagy, 1996). In the field of psychopathology, prediction has proved to be an elusive goal. Recently, there was an important multisite study that was performed in the USA in which three different treatment approaches to alcoholism were compared. An attempt was made to see whether particular groups of patients improved with specific treatments. The initial results have been negative. Except for differences related to the severity of patient symptomatology, other patient characteristics did not predict which treatments were most effective (Project Match Research Group, 1997).

The final goal of a classification is concept formation (Hempel, 1965). This goal is probably the most distant. In biological classification, Linneaus and his contemporaries made noteable gains in the classification of living organisms by creating classifications that served to describe most of the known information about these organisms. Almost a century later, Darwin formulated a theory of evolution which explained many of the organizational phenomena that Linneaus had used (Hull, 1988). In particular, Darwin's evolutionary theory provided a basis for understanding why there is a hierarchical arrangement about the categories in biological classification. The field of biological classification has continued to change. During the twentieth century, three different, competing approaches to biological classification have appeared (Hull, 1988). In oversimplified terms, one of these approaches emphasized the nomenclature and information retrieval purposes, the second focused on description, and the third was based on a theoretical view. The third view, the one based on theory, has become the dominant approach to biological classification (Nelson & Platnick, 1981).

4.03.3 DEVELOPMENT OF CLASSIFICATION SYSTEMS IN THE USA

The classification of mental disorders has an extensive history that can be traced back to some of the earliest writings known to man. A nineteenth century BC Egyptian writer discussed a disorder in women in which they would report vague and inconsistent physical symptoms that appeared to shift in body location over time (Veith, 1965). Psalm 102 provides an excellent clinical description of depression.

However, like many others areas of modern science, the first major commentaries on mental disorders were found in the writings of the Greeks. The Greek medical writers introduced four terms, all of which are still used today: "melancholia," "hysteria," "mania," and "paranoia." Melancholia was a Greek term that referred to a condition that now would be described by the word depression. Hippocrates believed that the sadness and the slowed bodily movements associated with this disorder were caused by an abundance of black bile, which he considered to be one of the four main ingredients in the human body. Hence, he named this disorder *melan* (black) + *cholia* (bile). The second term, hysteria, was the Greek word for the condition originally described by the Egyptians in which women had multiple, inconsistent and changing somatic complaints. The Hippocratic writers used the name of hysteria, which means pertaining to the uterus, because they believed that this disorder was caused by a dislodged, floating uterus. The last two terms were mania and paranoia. Mania, to the Greeks, referred to persons who were delusional. During the nineteenth century, individuals who were delusional but had few other symptoms were diagnosed with monomania (Berrios, 1993). Mania became an umbrella term for almost any type of psychotic state (Spitzka, 1883). The meaning of mania, however, changed again in the twentieth century to its contemporary denotation of grandiosity, excitement, expansiveness, and elation. The final Greek term, paranoia, has undergone a similar transformation. Paranoia initially meant craziness (*para* = abnormal + *nous* = mind). Now the term refers to people who are suspicious and often have delusions in which others are plotting against them.

After the Greeks, psychopathology did not attract much scientific interest until the nineteenth century. During the Middle Ages, mental disorders were associated with evilness. Thus, mental disorders were under the domain of religious authorities, not physicians or scientists. This approach to mental disorders began

to change in the late 1700s as exemplified by the fact that King George III of England, who was psychotic for most of the last decade of his reign, received medical care rather than religious counseling.

The first major American physician to be interested in mental disorders was Benjamin Rush who, as a signer of the Declaration of Independence, was one of the prominent American physicians of the late eighteenth century. He was very interested in the forms of insanity. Rush also published a book on the topic which he titled *Medical inquires and observations upon the diseases of the mind* (Alexander & Selesnick, 1966). He believed in a theory of neurosis. According to this theory, mental disorders were caused by overstimulation of the nervous system. Thus, environmental phenomena such as the pace of urban living, overuse of alcohol, excessive sexual behavior, masturbation, and smoking were all seen as casual factors in the development of mental disorders. As a result, asylums were the appropriate way to treat insane patients. Asylums could provide the quiet and tranquility that was necessary to allow the nervous system to heal and to repair itself.

About the same time that Rush was writing about psychopathology in the USA, there was an important discovery in France that was to markedly influence thinking about mental disorders. In 1822, a physician named Bayle (Quetel, 1990) performed autopsies on a number of patients who presented with gradiose delusions and dementia (i.e., who had lost their mind from the French *de-* (not) + *ment* (mind)). Bayle discovered that all of the patients in his study had marked changes in their brains. In addition to their dementia, all of these patients developed motor paralysis before they died. The brains of these patients had shrunk to almost half the weight of a normal brain, the skin of the brain (i.e., the meninges and the arachnoid) was thickened, and the color of the brain was strikingly different from that of normal brains. Bayle's name for this disorder was chronic arachnitis since he believed that this disorder was caused by a chronic infection of arachnoid tissue (Quetel, 1990). Later, the common name for this disorder was changed to dementia paralytica, a descriptive name that referred to the combined presence of a psychosis together with the progressive paralysis of the patient's limbs.

The discovery of dementia paralytica was the first instance in which a mental disorder had been shown to be associated with demonstrable changes in the brain. A number of autopsy studies appeared in the medical journals. These studies attempted to identify the exact neuropathology of this disorder as well as to understand its cause. Dementia paralytica was also a clinically serious disorder because it accounted for about one-sixth of all admissions to insane asylums during the nineteenth century. The prognosis of the disorder was very poor because death typically would occur within three years of the diagnosis (Austin, 1859/1976). For most of the nineteenth century, many different etiologies were proposed as potential causes of this disorder. Austin (1859/1976), for instance, listed the following as moral causes of dementia paralytica: death of a son, sudden loss of two good situations, wife's insanity, worry, and commercial ruin.

With the increasing interest in psychopathology during the nineteenth century, a number of classifications for mental disorders appeared. One example of these classifications was published by William A. Hammond (Blustein, 1991). Hammond, like Freud, was a nineteenth century neurologist. As a young physician, Hammond had published a set of interesting experimental studies on human physiology. At the age of 34, he became Surgeon General for the US Army during the Civil War and was credited with a number of important innovations at the time including hospital design, the development of an ambulance corps, and the removal of a mercury compound from the medical formulary. His political clashes with the Secretary of War, however, led to his court martial and dismissal. After the Civil War, he moved to New York City and developed a lucrative private practice as a neurologist—a remarkable accomplishment during a time when most physicians were generalists. His interests extended to psychiatry and to writing novels as well as to physiology, studies of sleep, hypnosis, and the use of animal hormonal extracts. Hammond wrote extensively in scientific journals. He was one of the founders of the *Journal of Nervous and Mental Disease* which is still published. In addition, he authored important textbooks of neurology and psychiatry.

In Hammond's textbook of mental disorders, he argued that there were six possible principles that could be used to organize a classification system:

(i) anatomical (organized by the part of the brain that is affected);

(ii) physiological (organized by the physiological system in the brain);

(iii) etiological (supposed causes);

(iv) psychological (based upon a functional view of the mind);

(v) pathological (observable, morbid alterations in the brain); and

(vi) clinical (descriptive, based upon clusters of symptoms).

Of these six principles, Hammond said that the anatomical, the physiological, and the pathological are the best, but he could not use them because the science of his time was insufficient. Hammond also rejected the etiological organization of categories, because he felt that an etiological classification, given nineteenth century knowledge, would be too speculative. Thus, the main choice was between the clinical (descriptive) approach and the psychological (mentalistic) approach. Hammond preferred the latter because he thought that a classification which did not have a strong theoretical basis would fail.

Hammond adopted a functional view of psychology that was common in his day. He believed that mental functioning could be organized into four areas: perception, cognition, affect, and will. Hence, he organized his classification of mental disorders into six major headings:

(i) perceptual insanities (e.g., hallucinations);

(ii) intellectual insanities (e.g., delusional disorders);

(iii) emotional insanities (e.g., melancholia);

(iv) volitional insanities (e.g., abulomania);

(v) compound insanities (i.e., disorders affecting more than one area of the mind); and

(vi) constitutional insanities (i.e., disorders with specific causes such as choreic insanity).

There were a total of 47 specific categories of mental disorders that were organized under these six major headings. Most names of these specific categories would not be recognized by modern clinicians. The descriptions of these disorders, together with case histories that he included in this textbook do, however, suggest that many of the disorders he was discussing would have modern counterparts. For instance, under the heading "intellectual insanities," Hammond classified four disorders whose names seem odd by modern standards: intellectual monomania with exaltation, chronic intellectual monomania, reasoning mania, and intellectual subjective morbid impulses. In modern terms, these disorders probably would be called biopolar I disorder (manic episode), schizophrenia (continuous), narcissistic personality disorder, and obsessive compulsive disorder.

In Hammond's textbook, the lengthiest discussion was devoted to general paralysis, for which Hammond's name was dementia paralytica. As part of this discussion, Hammond included the proclamation by John Michler (quoted at the beginning of this chapter). In his discussion of general paralysis, Hammond emphasized the many medical symptoms associated with this disorder.

By the end of the nineteenth century, there were three broad theories about the etiology of this disorder. First, one school of thought believed that it was caused by alcoholism because the disorder primarily affected men, the age of onset was typically during the 30s and 40s (which is the same time of onset for severe alcoholism), and most men with dementia paralytica had substantial drinking histories. Second, was the theory that the disorder was caused by syphilis. Epidemiological surveys had found that over 80% of men with dementia paralytica had had syphilis. However, since no survey had documented 100% with a history of syphilis, many investigators suggested that syphilis was an important precondition to the development of dementia paralytica but was not the single cause. Hammond, for instance, was clear that syphilis was not the etiology because syphilis was associated with other forms of insanity. The third theory was more psychological in that the disorder was believed to be caused by moral depravity because persons who drank, who frequented prostitutes, and who were in the military were more likely to have the disorder. As additional evidence, dementia paralytica was known to be rare among priests and Quakers.

Research, attempting to provide evidence for or against these theories, was performed. For instance, a famous German psychiatrist named Kraft-Ebbing performed a study in which he injected serum from patients with syphilis into the blood streams of patients with dementia paralytica. Since it was known at the time that a person could only develop a syphilitic infection once, if any of the patients with dementia paralytica developed syphilis, it would prove that they had not had syphilis previously. Hence syphilis could not be the cause of the disorder. None of the 32 patients with dementia paralytica developed syphilis. Kraft-Ebbing concluded that syphilis was the cause of this disorder.

The conclusive evidence regarding the etiology of dementia paralytica occurred in the early twentieth century. In 1906, the bacillus that causes syphilis was isolated, and the Waserman test for syphilis was developed. Plaut (1911) demonstrated that patients with dementia paralytica had positive Waserman tests from blood samples and also from samples of cerebrospinal fluid. In 1913, two Americans, Noguchi and Moore, were able to identify the presence of the syphillitic baccilus in the brains of patients who had died from dementia paralytica (Quetel, 1990). The name of this disorder was changed again to reflect the new understanding of this disorder. It was called paresis or general paralysis associated with tertiary syphilis of the

central nervous system. However, even after the discovery of the cause of dementia paralytica, the development of antibiotics to treat the disorder was another 30 years in the future. Thus, many patients were treated by inoculating them with malaria (Braslow, 1995).

4.03.4 KRAEPELIN

Another important development at the turn of the century was the international focus on the classificatory ideas of a German psychiatrist named Emil Kraepelin (Berrios & Hauser, 1988). Kraepelin was a researcher who initially learned his approach to research in a laboratory organized by Wundt, one of the founders of modern experimental psychology. After completing his medical degree, Kraepelin became the medical director for an insane asylum in east Prussia. While in that setting, Kraepelin published a number of experimental psychology studies on persons with mental disorders. He also began to write textbooks about psychopathology for German medical students. Like most textbook authors, Kraepelin organized his chapters around the major categories of mental disorders that he recognized. The sixth edition of Kraepelin's textbook (Kraepelin, 1902/1896) attracted major international attention because of two chapters in these texts.

One chapter was about a disorder that Kraepelin described as dementia praecox (praecox = early) which was a form of psychosis that had a typical age of onset in adolescence. Kraepelin recognized three subtypes to this disorder: hebephrenic, catatonic, and paranoid. Kraepelin's chapter on dementia praecox paralleled the immediately preceding chapter on dementia paralytica. Just as dementia praecox had three descriptive subtypes, dementia paralytica also had three descriptive subtypes: a depressed form, a grandiose form, and a paranoid form.

The second chapter to attract attention was what Kraepelin named manic-depressive insanity. Kraepelin's observations of patients in asylums had led him to believe that the mania and depression (= melancholia) had the same type of course when these patients were observed over time. Both were episodic disorders. Moreover, nineteenth century clinicians had recognized that there were some patients who went from episodes of mania to episodes of depression and vice versa. These observations led Kraepelin to hypothesize that mania and depression were essentially flip sides of the same coin. Hence, he combined, what had been recognized since the times of the ancient Greeks as two disorders, into one mental disorder.

In 1917, the newly formed American Psychiatric Association (APA) adopted a classification system that was quite similar to the classification contained in Kraepelin's sixth edition of his textbook. This early twentieth century American classification included the concepts of dementia praecox and manic-depressive disorder. The classification also adopted the fundamental Kraepelinian distinction between the organic disorders, the functional psychoses, and the neurotic/character disorders.

In 1932, the APA officially adopted a new classification system as part of the Standard Classified Nomenclature of Diseases (APA, 1933). This new classification, however, did not attract much attention (Menninger, 1963).

4.03.5 EARLY *DSMS* AND *ICDS*

World War II led to a renewed emphasis on classification. During the war, nearly 10% of all military discharges were for psychiatric reasons. By the time the war ended, there were four major competing classification systems at use in the USA: (i) the standard system adopted by the APA in 1932; (ii) the US Army classification; (iii) the US Navy classification; and (iv) the Veterans Administration system (Raines, 1952). In response to this disorganization, the APA formed a task force to create a system that would become standard in the USA. The result was the *Diagnostic and statistical manual of mental disorders* (*DSM*; APA, 1952). This classification is usually known by its abbreviated name: *DSM-I*. The *DSM-I* contained 108 different diagnostic categories.

The *DSM-I* was important for a number of reasons. First, the major rationale behind the *DSM-I* was to create a classification system that represented a consensus of contemporary thinking. Care was taken to include all diagnostic concepts that were popular in American psychiatry during the 1940s and 1950s. Thus, the *DSM-I* emphasized communication among professionals as the major purpose of classification and emphasized the need for psychiatric classification to be an accepted nomenclature that members of a profession can use to discuss their clinical cases. Consistent with this emphasis on communication, early versions of the *DSM-I* were revised, based on comments elicited in a questionnaire sent to 10% of the members of the APA. The *DSM-I* was finally adopted by a vote of the membership of the APA (Raines, 1952).

The emphasis on communication in the *DSM-I* led to a similar organizing movement at an international level. The international

psychiatric community had adopted a classification of mental disorders that was part of the *International statistical classification of disease, injuries and causes of death* (6th ed.) (*ICD-6* World Health Organization (WHO), 1948). The first *ICD* had been created in 1900 and was a medical classification of causes of death. The *ICD-6* was the first edition to include all diseases, whether they led to death or not. The classification of mental disorders in the *ICD-6* did not gain broad acceptance. A committee, chaired by the British psychiatrist Stengel, was formed to review the classification systems used by various countries and to make any necessary recommendation for changes to the WHO. What Stengel (1959) found was a hodgepodge of diagnostic systems between, and sometimes within, different countries. Stengel despaired over the confused state of international classification and said that the *ICD-6* did not serve as a useful nomenclature. A positive note in his review, however, concerned the *DSM-I*, which Stengel considered an advance over the other national classifications because of its emphasis on representing a well-organized nomenclature for a country.

As a result of Stengel's review, there was an international movement to create a consensual system that would be adopted by the WHO. The final product was the mental disorders section of the *ICD-8*. In the USA, the APA revised its *DSM* classification to correspond with the *ICD-8*. The US version of the *ICD-8* was known as the *DSM-II* (APA, 1968). The *DSM-II* had 185 categories. These categories were subdivided by a hierarchical organizational system. First, there was a distinction between psychotic and nonpsychotic disorders. The psychotic disorders were further subdivided into organic and nonorganic disorders. The classification of the organic disorders was in terms of etiology (e.g., tumors, infections, heredity, etc.). The nonorganic psychotic disorders primarily contained the Kraepelinian categories of schizophrenia and manic-depressive insanity. The nonpsychotic disorders were subdivided into eight subheadings including the neuroses (now called anxiety disorders), personality disorders, mental retardation, etc.

4.03.6 NEO-KRAEPELINIANS

After the publication of the *DSM-II*, psychiatric classification became a very unpopular topic. There were three general lines of criticism that were aimed at classification. First, the diagnosis of mental disorders was unreliable as shown by empirical research. Second, a number of critics attacked the implicit medical model approach to psychopathology that was associated with the *DSM-I* and *DSM-II*. Third, sociologists became interested in a theory of labeling that suggested the process of classification was a process that stigmatized human beings who adopted unusual patterns of behavior and the act of diagnosis could lead to self-fulfilling prophecies.

The first of these criticisms was summarized in three different review articles by Kreitman (1961), Zubin (1967), and Spitzer and Fleiss (1974). All discussed various problems associated with the classification of mental disorders and why poor reliability was a significant issue. Zubin (1967) made an excellent case that the lack of uniform statistical procedures for estimating reliability was a serious methodological problem. Spitzer and Fleiss (1974) suggested a solution to this problem and showed how this solution could be applied retrospectively to earlier data. Kreitman (1961) probably had the most far-reaching analysis of the issue because he said that the unreliability problem had been overemphasized and that the more serious issue was the unexplored validity of diagnostic concepts.

The second issue criticism of the early *DSM*s and *ICD*s was the implicit acceptance of a medical model. Despite the dramatic etiological solution to dementia paralytica, most of the twentieth century research has been disappointing to those who believed that mental disorders are caused by underlying biological processes. Large amounts of research have attempted to discover the etiology of disorders such as schizophrenia, yet, despite interesting advances, a clear understanding of the cause of this disorder is not available. A psychiatrist named Thomas Szasz published a book titled *The myth of mental illness* (Szasz, 1961). He argued that mental disorders are not diseases, but instead are better conceptualized as "problems in living." He argued that psychiatrists had placed themselves into the role of moral policeman to control individuals with deviant behavior patterns. Szasz is now considered one of a group of critics of classification known by the title "antipsychiatrists." Others in this group are the British psychiatrist Laing (1967), the French psychoanalyst Lacan (1977), and recent authors such as Sarbin (1997), and Kirk and Kutchins (1992).

The third criticism of classification that became popular in the 1960s and 1970s was the labeling criticism. Sociologists such as Matza (1969) and Goffman (1961) suggested that the act of psychiatric diagnosis could lead to self-fulfilling prophecies in which the behavior of deviant individuals was constrained to become even more deviant. A dramatic demonstration of

the labeling criticism was contained in a controversial paper published by Rosenhan (1973). In this paper, Rosenhan and his colleagues gained admission to mental hospitals even though they reported everything about themselves factually except their names and one auditory hallucination. All but one of these pseudopatients were admitted with a diagnosis of schizophrenia and all were released with a diagnosis of schizophrenia in remission. The pseudopatients commented that most of the patients were aware that they did not belong there, even though the hospital staff never figured that out. In addition, the experiences of the pseudopatients supported the labeling concern. For instance, one pseudopatient reported being bored while being on a ward and walking around. A nurse noticed him pacing and asked if he was feeling anxious. Following the publication of the Rosenhan paper, an issue of the *Journal of Abnormal Psychology* in 1975 was devoted to commentaries on this controversial paper.

Partially in reaction to these criticism of classification, a new school of thought was formed in psychiatry called the neo-Kraepelinians (Klerman, 1978). This group of psychiatrists, initially an active collection of psychiatric researchers at Washington University in St. Louis, believed that psychiatry, with its psychoanalytic emphasis, had drifted too far from its roots in medicine. The neo-Kraepelinians emphasized that psychiatry should be concerned with medical diseases, that extensive research was needed on the biological bases of psychopathology, and that much more emphasis needed to be placed upon classification if knowledge about psychopathology was to grow. Klerman (1978) summarized the perspective implicit in the neo-Kraepelinian approach to psychiatry by listing the following tenets:

(i) Psychiatry is a branch of medicine.

(ii) Psychiatry should utilize modern scientific methodologies and base its practice on scientific knowledge.

(iii) Psychiatry treats people who are sick and who require treatment.

(iv) There is a boundary between the normal and the sick.

(v) There are discrete mental illnesses. Mental illnesses are not myths. There is not one, but many mental illnesses. It is the task of scientific psychiatry, as of other medical specialties, to investigate the causes, diagnosis, and treatment of these mental illnesses.

(vi) The focus of psychiatric physicians should be particularly on the biological aspects of mental illness.

(vii) There should be an explicit and intentional concern with diagnosis and classification.

(viii) Diagnostic criteria should be codified, and a legitimate and valued area of research should be to validate such criteria by various techniques. Further, departments of psychiatry in medical schools should teach these criteria and not depreciate them, as has been the case for many years.

(ix) In research efforts directed at improving the reliability and validity of diagnosis and classification, statistical techniques should be utilized.

In 1972, a group of the psychiatric researchers at Washington University published a paper entitled "Diagnostic criteria for use in psychiatric research" (Feighner, Robins, Guze, Woodruff, Winokur, & Munoz, 1972). This paper listed 15 mental disorders that they believed had sufficient empirical evidence to establish their validity, and listed a set of diagnostic criteria for defining these disorders. The authors argued that a major problem in research about these disorders had stemmed from the lack of uniform definitions by different researchers of the disorders. They suggested that future research on any of these disorders should utilize the diagnostic criteria proposed in their paper.

The paper by Feighner et al. had a dramatic impact on American psychiatry. It was a heavily cited paper, probably the most frequently referenced journal article of the 1970s in psychiatry. The diagnostic criteria were immediately adopted and became the standard for psychiatric research. Moreover, the 15 categories described by Feighner et al. were expanded into a much larger set of categories, focusing primarily on the schizophrenic and affective disorders (Spitzer, Endicott, & Robins, 1975). This new classification was called the Research Diagnostic Criteria (RDC) and had an associated structured interview known as the SADS. Since the lead author of the RDC, Robert Spitzer, had been appointed as the new chairperson responsible for organizing the *DSM-III*, the RDC became the initial foundation from which the *DSM-III* developed.

4.03.7 *DSM-III*, *DSM-III-R*, AND *DSM-IV*

The *DSM-III* (APA, 1980) was a revolutionary classification. First, unlike the *DSM-I* and *DSM-II*, which had emphasized using consensus as the major organizing principle, the *DSM-III* attempted to be a classification based on scientific evidence rather than clinical consensus. For instance, the classification of depression was very different from the view of depression in the *DSM-I* and *DSM-II*, largely because of family history data gathered in research studies performed by the neo-Kraepelinians. In the earlier *DSM*s, the primary separation of affective disorders was in terms of a psychotic vs. neurotic distinction. The *DSM-III* dropped this differentiation and, instead, emphasized a separation of bipolar vs. unipolar mood disorders. Second, the *DSM-III* discontinued the use of prose definitions of the mental disorders. The neo-Kraepelinians were impressed by the research data suggesting that the reliability of psychiatric classification, as represented in the *DSM-I* and *DSM-II*, was less than optimal (Spitzer & Fleiss, 1974). To try to help improve diagnostic reliability, virtually all mental disorders in the *DSM-III* were defined using diagnostic criteria stimulated by the innovative system used in the Feighner et al. paper. Third, the *DSM-III* was a multiaxial classification. Because the *DSM-I* and *DSM-II* were "committee products," the subsections of these classifications had different implicit organizing principles. For instance, in the *DSM-I*/*DSM-II*, the organic brain syndromes were organized by etiology, the psychotic disorders were organized by syndromes, and the neurotic disorders were organized according to ideas from psychoanalytic theory. In order to avoid the confusion inherent in the use of multiple organizing principles, the *DSM-III* adopted a multiaxial system that permitted the separate description of the patient's psychopathology syndrome (Axis I), personality style (Axis II), medical etiology (Axis III), environmental factors (Axis IV), and role disturbances (Axis V). The *DSM-III* was published in 1980 and contained 265 mental disorders. Moreover, the size of the manuscript for the *DSM-III* was 482 pages, a huge increase over the 92 pages of the *DSM-II*.

The revolutionary impact of the *DSM-III* led to changes in many areas of the mental health professions. One area of impact was in terms of research. As soon as versions of the *DSM-III* began to be disseminated to researchers interested in mental disorders, new studies began to appear that explored the adequacy of the diagnostic criteria used in this classification. The *DSM-III* was a marked stimulus for descriptive research in psychiatry and in the other mental health professions. Another area of impact was political. There was a major controversy that erupted in the late 1970s over the issue of whether the term "neurosis" should appear in the *DSM-III*. Spitzer and the neo-Kraepelinians had exorcized this term from the classification because of its psychoanalytic associations. The psychoanalysts lobbied intensely within the APA to have the term reintroduced. Although a compromise was achieved (Bayer, 1981), the psychoanalysts lost ground in this struggle, and their influence in organized psychiatry has continued to wane since that time. A third area of impact for the *DSM-III* was economic. The *DSM-III* became very popular, sold well in the USA, and even became a surprisingly large seller to the international community as translations appeared. The sizeable revenues that accrued to the APA led to the formation of the American Psychiatric Press, which published subsequent versions of the *DSM* as well as many other books of clinical interest.

Despite its innovations and generally positive acceptance by mental health professionals, a number of criticisms were leveled at the *DSM-III*. One focus of criticism concerned the diagnostic criteria. Despite the intention to make decisions about the classificatory categories using scientific evidence, most diagnostic criteria were based on the intuitions of experts in the field. In addition, even though the goal when formulating diagnostic criteria was to make them as behavioral and explicit as possible, not all criteria met this goal. Consider part of the *DSM-III* criteria for histrionic personality disorder, for instance.

> Characteristic disturbances in interpersonal relationships as indicated by at least two of the following:
> 1) perceived by others as shallow and lacking genuineness, even if superficially warm and charming
> 2) egocentric, self-indulgent and inconsiderate of others
> 3) vain and demanding
> 4) dependent, helpless, constantly seeking reassurance
> 5) prone to manipulative suicidal threats, gestures or attempts

Note that common language terms such as "shallow" were highly subjective. In addition, a criterion such as 3) requires an inference about motivations and reasons for behavior, rather than direct observation of behaviors. Finally, subsequent research showed that criterion 5) above actually was observed more frequently in borderline rather than histrionic patients (Stangl, Pfohl, Zimmerman, Bowers, & Corenthal, 1985).

A second major criticism of the *DSM-III* concerned the multiaxial system. First, diagnosing multiple axes required increased time and effort by clinicians, an exercise they were unlikely to do unless they were certain that the gain in information was significant. Second, the relative emphasis on these five axes in the *DSM-III* was sizeably different. In the *DSM-III* manual, almost 300 pages were devoted to defining the Axis I disorders, another 39 pages were spent on Axis II disorders, whereas only two pages each were devoted to Axes IV and V.

Moreover, Axes I and II were assessed using diagnostic categories, whereas Axes IV and V were measured using relatively crude, ordinal rating scales. Third, the choice of particular axes was also criticized. For instance Rutter, Shaffer, and Shepard (1975) had advocated the use of one axis for the clinical syndrome in childhood disorders with a second focusing on intellectual level. Instead, both clinical syndromes and mental retardation were Axis I disorders in the *DSM-III*. A group of psychoanalysts argued that defense mechanisms should have been included as a sixth axis. Psychiatric nurses advocated an additional axis for a nursing diagnosis, relevant to the level of care required by a patient.

Only seven years later, a revision to the DSM-III was published. This version, known as the *DSM-III-R* (APA, 1987), was intended primarily to update diagnostic criteria using the research that had been stimulated by the *DSM-III*. It was called a revision because the goal was to keep the changes small. However, the differences between the *DSM-III* and the *DSM-III-R* were substantial. Changes included renaming some categories (e.g., paranoid disorder was renamed delusional disorder), changes in specific criteria for disorders (e.g., the criteria for schizoaffective disorder), and reorganization of categories (e.g., panic disorder and agoraphobia were explicitly linked). In addition, six diagnostic categories originally in the *DSM-III* were deleted (e.g., egodystonic homosexuality and attention deficit disorder without hyperactivity) while a number of new specific disorders were added (e.g., body dysmorphic disorder and trichotillomania). As a result, the *DSM-III-R* contained 297 categories compared to the 264 categories in the *DSM-III*.

Associated with the *DSM-III-R* was the development of a major controversy that had political overtones. Among the changes proposed for the *DSM-III-R* was the addition of three new disorders: premenstrual syndrome (PMS), masochistic personality disorder and paraphilic rapism. These additions raised the ire of a number of groups, especially feminists. Concerning PMS, feminists argued that the inclusion of this disorder into the *DSM* would be the implicit assumption that the emotional state of women can be blamed on their biology. If it were to be a disorder, the feminists argued that PMS should be classified as a gynecological disorder rather than a psychiatric disorder. Masochistic personality disorder had been suggested for inclusion by psychoanalysts who pointed to the extensive literature on this category. Feminists, however, believed that this diagnosis would be assigned to women who had been physically or sexually abused. Thus, this diagnosis would have the unfortunate consequence of blaming these women for their roles as victims. Finally, the proposal to include paraphilic rapism was also attacked. The critics argued that this diagnosis would allow chronic rapists to escape punishment for their crimes because their behaviors could be attributed to a mental disorder. Thus, these men would not be held responsible for their behaviors.

Because of the ensuing controversy, a compromise somewhat similar to the earlier compromises about homosexuality and neurosis was attempted. The authors of the *DSM-III-R* revised the names of the first two disorders (PMS and masochistic personality disorder) to periluteal phase dysphoric disorder and self-defeating personality disorder. They also deleted the proposal to add paraphilic rapism. In addition, another disorder, sadistic personality disorder, was added presumably to blame abuser as well as victims, thereby balancing the potential antifeminine connotations of self-defeating/masochistic personality disorder. This compromise was not successful. As a result, the executive committee for the American Psychiatric Classification decided not to include these categories in the body of the *DSM-III-R*. Instead, they were placed in an appendix as disorders needing more research (Walker, 1987).

The *DSM-IV* was published in 1994, contained 354 categories, and was 886 pages in length, a 60% increase over the *DSM-III* and almost seven times longer than the *DSM-II* (APA, 1994). There are 17 major categories in the *DSM-IV*:

disorders usually first diagnosed in childhood
cognitive disorders
mental disorders due to a general medical condition
substance-related disorders
schizophrenia and other psychotic disorders
mood disorders
anxiety disorders
somatoform disorders
factitious disorders
dissociative disorders
sexual disorders
eating disorders
sleep disorders
impulse control disorders
adjustment disorders
personality disorders
other conditions that may be a focus of clinical attention

The *DSM-IV* retained a multiaxial system and recognized five axes (dimensions) along which patient conditions should be coded:

Axis I clinical disorders
Axis II personality disorders/mental retardation
Axis III general medical conditions
Axis IV psychosocial and environmental problems
Axis V global assessment of functioning

As with the *DSM-III-R*, a major focus in the *DSM-IV* revision concerned diagnostic criteria. A total of 201 specific diagnoses in the *DSM-IV* were defined using diagnostic criteria. The average number of criteria per diagnosis was almost eight. Using this estimate, the *DSM-IV* contains slightly over 1500 diagnostic criteria for the 201 diagnoses.

To give the reader a glimpse of how the diagnostic criteria have changed from the *DSM-III* to the *DSM-IV*, the criteria for histrionic personality disorder are listed below:

A pervasive pattern of excessive emotionality and attention seeking, beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following:
(1) is uncomfortable in situations in which he or she is not the center of attention
(2) interaction with others is often characterized by inappropriate sexually seductive or provocative behavior
(3) displays rapidly shifting and shallow expression of emotions
(4) consistently uses physical appearance to draw attention to oneself
(5) has a style of speech that is excessively impressionistic and lacking in detail
(6) shows self-dramatization, theatricality, and exaggerated expression of emotion
(7) is suggestible, i.e., easily influenced by others or circumstances
(8) considers relationships to be more intimate than they actually are

In addition to presenting diagnostic criteria, the *DSM-IV* contains supplementary information about the mental disorders in its system. For instance, there are three pages of information about histrionic personality disorder including

diagnostic features (prose description of symptoms)
associated features and disorders (mental disorders that are likely to co-occur)
specific culture, age and gender features
prevalence
differential diagnosis (how to differentiate the disorder from others with which it is likely to be confused)

In order to help ensure that the *DSM-IV* would be the best possible classification of mental disorders, the steering committee for this classification contained 27 members, including four psychologists. Reporting to this committee were 13 work groups composed of 5–16 members. Each work group had a large number of advisors (typically over 20 per work group).

There were three major steps associated with the activities of each work group. First, all work groups performed extensive literature reviews of the disorders under their responsibility. Many of these literature reviews were published in the journal literature. Second, the work groups solicited contributions of descriptive data from researchers in the field. Using these data, the work groups reanalyzed the data to decide which diagnostic criteria needed revision. Third, a series of field trials was performed on specific topics. For instance, the personality disorders work group performed a multicenter study on antisocial personality disorder which led to a significant alteration in the diagnostic criteria for that disorder.

The *DSM-IV* was not without controversy. For instance, the issues that had been raised in the *DSM-III-R* regarding premenstrual syndrome, masochistic personality disorder, and sadistic personality disorder continued in the *DSM-IV*. Interestingly, none of these three disorders were included in the *DSM-IV*. In fact, two (masochistic and sadistic personality disorders) completely disappeared from the classification. PMS remained in an appendix as a disorder "for further study." Interestingly, 17 other disorders joined PMS in this appendix as did three possible new axes for the multiaxial system (defense mechanisms, interpersonal functioning, and occupational functioning).

Earlier editions of the *DSM* had few, if any, references to document the sources for any factual claims in these classifications. The *DSM-IV* attempt to overcome this problem was by publishing a five-volume companion set of sourcebooks. These sourcebooks are edited papers by members of the work groups. The intent of the sourcebooks is to provide a scholarly basis for understanding the specific decisions that were made by the work groups.

4.03.8 *ICD-9, ICD-9-CM,* AND *ICD-10*

Earlier in this chapter, the point was made that the *DSM-II* and *ICD-8* were virtually identical because the American psychiatric community had joined an international movement to create a consensual classification. With the revolutionary *DSM-III*, American psychiatry reversed itself and created a radically new classification based upon the purpose of

description, rather than emphasizing a system that would be acceptable world-wide.

The editions of the *ICD*s were intended to be revised every 10 years. The *ICD-8* was published in 1966; the *ICD-9* came out in 1977. The mental disorders section of the *ICD-9* was very similar to the *ICD-8/DSM-II* (WHO, 1978). The psychotic/nonpsychotic distinction was the primary hierarchical distinction among categories. The psychotic disorders were further subdivided into organic and functional psychoses. There were 215 categories in this system, and the *ICD-9* was published in a monograph that was 95 pages in length.

The USA has signed an international treaty that obliges it to use the *ICD* as the official medical classification. Thus, when the *DSM-III* was created, an odd numeric coding scheme was incorporated so that the *DSM-III* categories could be incorporated with the *ICD-9* framework. To understand this, below is an overview of the specific diagnostic categories under the general heading of anxiety disorders in the *DSM-III*:

Phobic disorders
300.21 Agoraphobia with panic attacks
300.22 Agoraphobia without panic attacks
300.23 Social phobia
300.29 Simple phobia
Anxiety states
300.01 Panic disorder
300.02 Generalized anxiety disorder
300.30 Obsessive compulsive disorder
Post-traumatic stress disorder
308.30 Acute
309.81 Chronic or delayed
300.00 Atypical anxiety disorder

Notice that the coding scheme for the anxiety disorders in the *DSM-III* is not what one might expect. Most of the anxiety disorders are coded with 300.xx numbers. However, the two forms of post-traumatic stress disorder are coded 308.30 and 309.81. Notice also that the first number after the decimal point is somewhat irregular. The phobic disorders, listed first in the *DSM-III*, are given 300.2x codes whereas the anxiety states are coded 300.0x or 300.30.

To understand why the *DSM-III* codes appear this way, below is a listing of the specific neurotic disorders in the *ICD-9*:

Mental disorders (290–319)
Nonpsychotic mental disorders (300–316)
Neurotic disorders (300)
300.0 Anxiety states
300.1 Hysteria
300.2 Phobic states
300.3 Obsessive-compulsive disorders
300.4 Neurotic depression
300.5 Neurasthenia
300.6 Depersonalization syndrome
300.7 Hypochondriasis
300.8 Other neurotic disorders
300.9 Unspecified

In the *ICD-9* system, all diagnoses have four-digit codes. The codes for all mental disorders range from 290 to 319. The 29x disorders are the psychotic disorders; 300–315 are reserved for nonpsychotic disorders; and 316–318 are codes for classifying mental retardation. The first subheading under the nonpsychotic disorder is the neurotic disorders. Notice that this subheading includes what the *DSM-III* recognizes as anxiety disorders but it also includes categories that the *DSM-III* placed under other headings (e.g., neurotic depression = dysthymic disorder and depersonalization syndrome). Because the *DSM-III* anxiety disorders were mostly found under the *ICD-9* neurotic disorders, these categories were given 300.xx codes. However, post-traumatic stress disorder (chronic; PTSD) was given a code number of 309.89 because it was included under the general *ICD-9* heading of adjustment reactions. Notice also that PTSD has an xxx.8 coding. In the *ICD-9* coding system, all diagnoses with an xxx.8 code represent country-specific categories that are generally recognized by the international psychiatric community. Thus, PTSD (309.89) is a US category that has no clear equivalent in the international diagnostic system. Another *DSM-III* category with a coding that reflects a similar status is borderline personality disorder (301.83).

In order to blend the *DSM-III* with the *ICD-9* so that a consistent coding system would be used, a US version of the *ICD-9* was created. This new version was the *ICD-9-CM* (where CM stands for "clinical manual"). The *ICD-9-CM* is the official coding system that all physicians and mental health professionals must use when assigning diagnostic codes. However, American clinicians do not need to refer to the *ICD-9-CM* because the applicable codes are listed in the printed versions of the *DSM-III*, *DSM-III-R* and *DSM-IV*.

As noted earlier, the *DSM-III* and its successors (*DSM-III-R* and *DSM-IV*) were resounding successes. Not only did these systems become dominant in the USA, but they also achieved substantial popularity among European mental health professionals (Mezzich, Fabrega, Mezzich, & Coffman, 1985). The proponents of the *ICD* were somewhat resentful (Kendell, 1991). Thus, when the *ICD-10* was created, substantial changes were made that utilized innovations from the *DSM-III*.

First, like the *DSM-III*, the *ICD-10* went through extensive field testing (Regier, Kłaeber, Roper, Rae, & Sartorius, 1994). There were two major goals in the field testing: (i) to ensure that the *ICD-10* could be used in a reliable way across clinicians, and (ii) to examine the acceptability of the mental disorder categories contained in this system. The data reported regarding both of these goals have given a favorable view of the *ICD-10*.

The second important innovation of the *ICD-10* is that the mental disorders section is published in two forms. One form, the blue manual, is subtitled *Clinical descriptions and diagnostic guidelines* (WHO, 1992). The blue manual contains prose definitions of categories and is primarily intended for clinical use. The other, the green manual, is like the *DSM-III* in that the categories are defined using explicit diagnostic criteria with rules regarding how many criteria must be met in order for a diagnosis to be made (WHO, 1993). The green manual is intended for research use.

The complete *ICD-10* is organized into a series of 21 chapters, one of which is Chapter V (labeled with the prefix F) about "Mental and behavioural disorders." Other chapters in the *ICD-10* are:

Chapter I A–B Infections and parasitic diseases

Chapter II C–D Neoplasms

Chapter X J Diseases of the respiratory system

Chapter XXI Z Factors influencing health status and contact with health services

In terms of classificatory size, the mental disorders section of the *ICD-10* and the *DSM-IV* are reasonably similar. The *DSM-IV* contains 354 categories organized under 17 major headings. The *ICD-10* has 389 categories that are structured under 10 major headings. One ironic feature of the *ICD-10* is that it did not adopt a multiaxial system of classification. This decision is ironic because the idea originated in Europe, most prominently by a Swedish psychiatrist, Essen-Moller (1971).

4.03.9 CONTROVERSIES AND ISSUES

Although the *DSM-III* and its successors are usually viewed as substantial improvements in the classification of mental disorders, a number of controversies and issues still remain concerning the classification of mental disorders. The remainder of this chapter attempts to provide an overview of some of these issues.

4.03.9.1 Organizational Models of Classification

There are four organizational models of classification that have been frequently discussed in the literature. Often these models are seen as competing and incompatible. For instance, most discussions of the first two models, the categorical and the dimensional, are presented as if the mental health professions must choose between them. However, hybrid models are possible (Skinner, 1986) and perhaps even the most probable.

Mental disorders are usually discussed as if they are categories. Thus, a categorical model is often implicitly assumed to be the structural model for psychopathology (Kendell, 1975). The tenets of this model are listed below:

1.1 The basic objects of a psychiatric classification are patients
1.2 Categories should be discrete, in the sense that the conditions for membership should be able to be clearly ascertained
1.3 Patients either belong or do not belong to specified classes (categories)
1.4 The members of a category should be relatively homogeneous
1.5 Categories may or may not overlap
1.6 In the borderline areas where categories may overlap, the number of overlapping patients should be relatively small
1.7 Cluster analytic methods can be used to initially identify categories
Discriminant analysis is used to validate categories. (Blashfield, 1991, p. 14)

The *DSM*s, particularly the *DSM-III* and its successors, are seen as fitting a categorical model. According to the categorical model, the unit of analysis is the patient. Diagnoses refer to classes of patients. Patients either are or are not the members of the categories. The categorical model assumes that some type of definitional rule exists by which the membership in a category can be determined. Moreover, membership in a category is considered to be a discrete, all-or-nothing event. An animal is either a cat or not a cat. A patient is either a schizophrenic or not a schizophrenic.

An important assumption of the categorical model is that members of a category are relatively homogeneous. All animals that belong to the class of "birds" should be reasonably similar morphologically. This is not to say that all birds must be alike. Certainly a robin and sparrow have a number of obvious differences. Yet they are more like each other than either is to a lynx, for instance. In the same way, two schizophrenic patients should be relatively similar. Both may have different symptom

pictures, but their symptom pictures should be more similar to each other than either is to an antisocial patient (Lorr, 1966).

Classes in a categorical model may or may not overlap. Most uses of the categorical model typically assume that nonoverlapping categories are an ideal condition, but recognize that this condition may not always happen. Thus, overlap among categories is often treated like measurement error—a condition to be tolerated, but minimized. However, there are categorical models that have been developed in which the categories are assumed to overlap (Jardine & Sibson, 1971; Zadeh, 1965). In these models, overlap is not error. Categories are fuzzy sets whose boundaries of membership do not need to result in mutually exclusive groupings.

According to the assumption of relative homogeneity, the number of patients who clearly belong to one and only one category should be relatively frequent, whereas patients who fall in the overlapping, borderline areas between categories should be relatively infrequent. A sparrow–lynx should not occur if categories are to have the necessary homogeneity that allows them to be separable constructs (Grinker, Werble, & Drye, 1968).

Finally, the methods that have been developed to find the boundaries among categories are called cluster analytic methods (Everitt, 1974). Generally, these methods analyze a large matrix of descriptive data on patients, and attempt to form clusters (categories) of relatively homogeneous patients in terms of the descriptive variables that were gathered on the patients. Cluster analysis was used in the 1960s and 1970s to create new descriptive classifications. In the last decade, most researchers have abandoned the use of these methods because of pragmatic difficulties in their application and because of unsolved statistical issues. Meehl, however, has developed a related method for isolating categories that he believes has promise (Meehl, 1995).

Although the categorical model, as presented above, seems to be a common sense model of psychiatric classification, the recent *DSM*s clearly do not adhere to this model. First, as noted above, a categorical model assumes that the unit of analysis is the patient and that groups of patients are members of similar sets called mental disorders. The authors of the *DSM-III, DSM-III-R and DSM-IV* explicitly reject this approach. They state that these classifications are not intended to classify individual patients (Spitzer & Williams, 1987). Instead, these recent *DSM*s state that they are classifying disorders (rather than patients).

A second structural model is the dimensional model. The tenets for this model are:

2.1 The basic unit of the dimensional model is a descriptive variable (e.g., a symptom, a scale from a self-report test, a laboratory value, etc.)
2.2 Dimensions refer to higher-order, abstract variables
2.3 A dimension refers to a set of correlated descriptive variables
2.4 There are a relatively small number of dimensions compared to the number of descriptive variables, yet the dimensions account for almost as much reliable variance as do the larger number of descriptive variables
2.5 Dimensions themselves may be correlated or independent
2.6 The methods used to identify dimensions are exploratory factor analysis and multidimensional scaling. Confirmatory factor analysis can be used to test a specific dimensional model. (Blashfield, 1991, p. 15)

For the dimensional model, the basic units of analysis are the descriptive variables. Thus, the dimensional model focuses on symptoms, behaviors, diagnostic criteria, scales from self-report tests, and the like. The dimensional model summarizes these descriptive variables by forming higher-order abstract variables that can serve to represent the original measurement variables. Each of these higher-order abstract variables constitutes a dimension through its conceptualization as occurring on a continua. Patients can have scores anywhere along these dimensions.

A major test of a dimensional model is parsimony. The specific dimensions in such a system should account for most of the systematic, reliable variance that exists within the original set of descriptive variables. If the dimensions do not account for the reliable variance in the original descriptive variables, then using the dimensions will sacrifice a great deal of information and the original variables should be used rather than the smaller set of dimensions.

A third structural model that is often discussed regarding the classification of mental disorders is a disease model. The basic assumption in this model is that all diagnostic categories refer to medical diseases (Wing, 1978). In effect, this model is a modern extension of Griesinger's famous nineteenth-century dictum that all mental disorders are diseases of the brain (Stromgren, 1991). The tenets of the disease model are:

3.1 The fundamental units are biological diseases of individual patients (essentialism)
3.2 Each diagnosis refers to a discrete disease
3.3 Diagnostic algorithms specify objective rules for combining symptoms to reach a diagnosis
3.4 Adequate reliability is necessary before any type of validity can be established
3.5 Good internal validity will show that the category refers clearly described patterns of symptoms
3.6 Good external validity will mean that the diagnosis can be used to predict prognosis, course and treatment response. (Blashfield, 1991, p. 8)

Some authors have assumed that a categorical model and a disease model are the same. These models are not identical. A categorical is neutral about the existential status of the categories in its system. A disease model adopts a stronger view. Diseases do have an existential status. Diseases are real. The goal of medical research is to identify, describe, understand and eventually treat these diseases. This belief in the reality of diseases is associated with a broader view about the status of scientific concepts known as essentialism.

Notice also that diseases are not necessarily categorical, at least as this model was described above. For instance, more than one disease can occur in the same patient. In fact, some diseases are very likely to co-occur (e.g., certain sarcomas have high frequency in patients with AIDS). Thus, diseases do not refer to mutually exclusive sets. In addition, there are diseases that are conceptualized as dimensional constructs. Hypertension is the most common example. Patients with hypertension vary along a continuum. A categorical scaling of this continuum is possible, but imposing a categorical separation on this continuum is arbitrary.

The fourth model of classification is the prototype model. Cantor, Smith, French, and Mezzich (1980) have suggested that this model is superior to the implicit categorical model of psychiatric classification. For those readers who do not know what the prototype model is, the easiest way to conceptualize this model is through an example.

According to the prototype model, if a mother wanted to teach a child what "anger" means, she would not say "Steven, you need to understand that anger is an emotion that many of us feel when we are frustrated or upset." Instead, at some point when little Steven is upset because he has to go to bed and he tries to hit his mother, she would say "I know that you are angry, Steven, but it is time to go to bed." And on another day, when Steven is upset because another child took one of his toys, his mother might say, "You are feeling angry, Steven. Being angry is natural, but you should not hit the other child. Maybe if you ask Carol she will return your toy."

In effect, a child learns a concept by being presented with instances of the concept. Once the child is able to associate the instances with a verbal label (i.e., the word "angry"), then the child will have a grasp of the concept. Later, the child will learn to abstract the essential features of the concept. This occurs by making observations about similarities that occur among instances of the concept (e.g., internal feelings, interpersonal context, etc.). Russell and Fehr (1994) provide an interesting and more complete discussion of the concept of anger from a prototype perspective.

Another important aspect of the prototype model is the idea that not all instances of a concept are equally good representatives of the concept. A robin, for instance, is a good exemplar of a bird. Robins are about the same size as most birds; they have feathers; they can fly; etc. Penguins, however, are not a good exemplar. Penguins are larger than most birds; they cannot fly; they do have feathers, although, to a child, their covering probably seems more like fur than feathers; etc.

The above presentation of the prototype model is easy to understand and seems like a common-sense view of classification. However, advocates of the prototype model argue that this model is radically different than a categorical model (Barasalou, 1992; Russell, 1991). According to the categorical model, classificatory concepts are defined by listing the features that are sufficient for making a diagnosis. If a given instance has a sufficient number of these features, then that instance is a member of the classificatory concept. Moreover, all members of a concept are equal. A square is a square. One square is not squarer than another square. In contrast, the prototype model does stipulate that some instances of a concept are better exemplars than others. The Glenn Close character in *Fatal attraction* is a better representation of borderline personality disorder than the Jessica Walter character in *Play Misty for me*.

The basic tenets of the prototype model are presented below:

4.1 Diagnoses are concepts that mental health professionals use (nominalism)
4.2 Categories are not discrete
4.3 There is a graded membership among different instances of a concept
4.4 Categories are defined by exemplars
4.5 Features (symptoms) are neither necessary nor sufficient to determine membership
4.6 Membership in a category is correlated with number of features (symptoms) that a patient has. (Blashfield, 1991, p. 11)

The major difference between the disease and the prototype model is that the latter is associated with nominalism. Nominalism is the position that the names of diseases are just convenient fictions that clinicians use to organize information. Diagnostic terms do not have

some underlying reality of their own. Diagnostic concepts are names. These concepts exist simply to meet some functional goal.

The preceding discussion of organizational models of the classification of psychopathology is overly simplistic. Each of these models, when examined more closely, can become quite complex, and the apparent distinctions between the models blur. Two instances of this complexity will be described. First, although the categorical and dimensional are usually presented as if they are competing and antagonistic, Skinner (1986) has suggested that these models are actually complementary. He suggested that the measurement model associated with a dimensional perspective is the more fundamental of the two models. The dimensional model only assumes that, in order to assess a patient, a clinician should gather information on specific descriptive variables that are correlated and that can be summarized by higher-order variables (dimensions). The categorical model also assumes that descriptive variables can be sorted into dimensions. But, in addition, the categorical model asserts that the patients themselves "cluster" into groups that are relatively homogeneous on these dimensions. Thus, from Skinner's hybrid perspective, a pure categorical model makes stronger assumptions about descriptive data than does a dimensional model. However, psychological models of human social cognition suggest that categorical models are more basic than are dimensional models (Wyer & Srull, 1989).

A second example of the complexity of these models is associated with Barsalou's distinction between "prototype" and "exemplar" models. There are two types of approach that can be used to define a concept: intensional definitions and extensional definitions. An intensional definition lists the features that can be used to separate the concept from related concepts (e.g., a square is a four-sided closed figure whose sides are equally long and occur at right angles to each other). In contrast, an extensional definition is a definition by listing the members of the category (e.g., the 1957 New York Yankees included Roger Marris, Yogi Berra, Micky Mantle, etc.). Barsalou says that a prototype model uses an intensional definition for categories in which the prototype represents the average (centroid) of the concept. Using the example of a child learning about birds, Barasalou suggests that the reason that a robin is a prototype for bird, whereas a penguin is not, is that robins have the average features of most birds (small size, bright coloring, migration, food choices, etc.). Penguins, in contrast, are statistically unusual on these dimensions. An exemplar model, according to Barsalou, uses a particular type of extensional definition in which a concept is defined by an outstanding or exemplary instance of the concept. Thus, Micky Mantle might be a good exemplar of the 1957 New York Yankees even though Mantle's batting prowess was hardly average. In the same way, Abraham Lincoln might be seen as exemplar of American presidents, even though he was not average for this set of individuals.

4.03.9.2 Concept of Disease

The discussion of the disease model vs. the prototype model led to a brief introduction regarding the dualism of essentialism vs. nominalism. This dualism is associated with a complicated problem that has bothered writers about classification throughout the last two centuries. What is a disease? Do the two concepts of "mental disease" and "mental disorder" have the same meaning? To discuss the issues associated with the meaning of "disease," the writings of a British internist named Scadding will be discussed. At the end of this section, other approaches to the concepts of "disease" and "disorder" are briefly introduced.

Scadding's (1959) first attempt to discuss the meaning of disease occurred in a short essay. This essay offered his first general definition of disease which read:

> The term "a disease" refers to those abnormal phenomena which are common to a group of living organisms with disturbed structure or function, the group being defined in the same way.

In effect, Scadding was saying that a disease was associated with a cluster of signs and symptoms (i.e., abnormal phenomena) that are associated with some functional or structural disturbance in the human body. Scadding went on to argue that a disease had (i) defining characteristics and (ii) diagnostic criteria. The defining characteristics refer to the indications that prove the presence of the disease (e.g., locating syphilitic bacilli in the brains of individuals with paresis). In contrast, the diagnostic criteria are signs and symptoms of the disease that may or may not be present (e.g., motor paralysis, grandiose delusions, and sluggish pupillary response to light would be possible diagnostic criteria for paresis).

Ten years later, Scadding (1969) revised his definition of disease to read as follows:

> A disease is the sum of the abnormal phenomena displayed by a group of living organisms in association with a specified common characteristic or set of characteristics by which they differ from the norm for their species in such a way as to place them at biological disadvantage.

There are four important points to note about this second definition of disease. First, the emphasis is on abnormal phenomena. Scadding wanted to be quite clear that the name of the disease does not refer to the etiologic agent causing the disease. That is, tuberculosis is not simply defined by the presence of a particular bacterium, *Mycobacterium tuberculosis*. To have tuberculosis a patient must manifest the symptoms of the disease as well as the anatomical changes (i.e., the formation of characteristic lesions called tubercles in the lung) associated with this disease. This distinction is important because there are other bacilli, besides *Mycobacterium tuberculosis*, which can cause these lesions and the symptom pattern of tuberculosis.

Second, the definition contains the rather vague phrase "common characteristic." Scadding argued that there are three general ways of characterizing any individual disease: (i) a clinical-descriptive approach, (ii) a morbid anatomical approach, and (iii) an etiological approach. (Note that these approaches were presented almost a century earlier by Hammond (1883), as previously discussed.) The clinical-descriptive approach is simply the description of the "syndrome." That is, the clinical-descriptive approach outlines a loose cluster of signs and symptoms that are correlated in their appearance in patients. For instance, the clinical-descriptive approach to defining diabetes focuses on frequent urination, an intense thirst, and rapid loss of weight as indications of this disorder. The clinical-descriptive approach dominated when the *DSM-III* and its successors were created. The second approach concerns morbid anatomy. This refers to the anatomical changes in the body's structure associated with the disease. For diabetes mellitus, a morbid anatomy view might define this disease in terms of the destruction of β-insulin-producing cells in the pancreas. Finally, the etiological approach would be to define a disease in terms of the syndrome caused by a known and specifiable etiological process. For Type I diabetes mellitus, this might be an autoimmune process whose exact details have yet to be specified. For paresis, the etiological agent is the effect of the syphilitic bacillus on the central nervous system of the affected individual.

Scadding commented that, historically, knowledge about diseases typically proceeds from clinical-description to morbid anatomy to etiology. Certainly his observation seems to be correct when applied to the history of paresis. He argued that any of these approaches to characterizing a disease are appropriate. That is, a disease can be defined in terms of a clinical syndrome; or it can be defined by some associated morbid anatomy; or it can be defined through a recognition of its etiological cause.

The third point to note about Scadding's definition is its emphasis on norms, in that disease refers to an abnormality. To be a disease, the condition must refer to phenomena that are statistically deviant. For instance, most of us have various types of bacteria that normally reside in our intestines and which are important in the digestive process. The presence of these bacteria do not define a disease. The effects of these bacteria are normative. In fact, individuals without these bacteria have abnormal digestive processes.

Finally, the definition ends with the term "biological disadvantage." Scadding introduced this term because he recognized that not all nonnormative aspects of human structure and functioning should be called diseases. For instance, some individuals produce an abnormal amount of ear wax. However, this should not define a disease unless there is some biological disadvantage associated with this condition. Although the term biological disadvantage is not more precisely specified, its general meaning seems clear: syphilis and diabetes place an individual at biological disadvantage since both can lead to death if untreated.

In 1979, Scadding and two Canadian authors (Campbell, Scadding and Roberts, 1979) extended their ideas about disease by studying what physicians and nonphysicians meant by the concept of disease. They published a survey that they had conducted regarding the meaning of disease. To conduct their survey, these authors read a list of possible diseases to four groups of individuals: (i) a group of medical faculty, (ii) a group of nonmedical faculty, (iii) a sample of general practice physicians, and (iv) a sample of youth in British and Canadian schools. The subjects in this study were asked to note whether the terms being read aloud referred to diseases or not. In addition, the subjects were asked to assign degree of confidence ratings to their decisions.

At the top of the list of conditions that were viewed as diseases are infections (malaria, tuberculosis, syphilis, measles). Virtually everyone in the four groups, whether physicians or nonphysicians, agreed that these terms referred to diseases. Syphilis, for instance, was considered a disease by over 90% of the subjects in all groups.

At the bottom of the list were concepts that were not considered diseases by these subjects. Two terms that were seen as referring to diseases by less than 30% of all four groups were drowning and starvation. Many of the terms at the bottom of Scadding's list might be described

as injuries, i.e., traumas that affect bodily functioning and that were caused by identifiable external events such as a car accident (e.g., skull fracture) or ingestion of a substance (e.g. barbituate overdose, poisoning).

The psychiatric concepts in the list (schizophrenia, depression, and alcoholism) were ranked in the middle. There was considerable variance among the four groups regarding whether these concepts referred to diseases. For instance, faculty of medical schools rated these three concepts in the following order: schizophrenia (78%), depression (65%), and alcoholism (62%). Children in secondary schools had quite different impressions of what is considered to be a disease: alcoholism (76%), schizophrenia (51%) and depression (23%).

One factor that had a large influence on whether a term referred to a disease concerned the role of a physician in the diagnosis or treatment of the disorder. Malaria and syphilis require a doctor to diagnose and treat. In contrast, starvation can be identified and treated by nonmedical individuals. The latter is also true of acne and hemorrhoids, although the intervention of physicians can prove useful for both. Consistent with this view, acne and hemorrhoids were ranked in the middle of the list. The potential role of nonphysicians in the treatment of mental disorders may also account for the occurrence of schizophrenia, depression and alcoholism in the middle of the same list. Scadding et al (1979, p. 760) concluded their paper with the following interesting comment:

> Most people without medical training seem to think of a disease as an agent causing illness. The common concept of "disease" is essentialist: diseases exist, each causing a particular sort of illness. Doctors tend to adopt a more nominalist position, but they obviously retain remnants of belief in the real existence of diseases.

When viewed from this dualism of an essentialist vs. a nominalist perspective, Scadding had started his definitional attempts from an essentialist perspective but, by the time of his last writings on the topic, he was suggesting that a nominalist view was preferred. Interestingly, the writings of a prominent British psychiatrist, Kendell, who has also tried to solve this issue, have followed the same progression. His ideas shifted from a paper trying to settle on an essentialist meaning of disease (Kendell, 1976) to a skeptical discussion of how the disease model fails to explain alcoholism (Kendell, 1979) to a nominalist view (Kendell, 1986).

Because this nominalism vs. essentialism dualism is so important, the approach of Wulff, Pedersen, and Rosenberg (1986) is discussed briefly. To understand the dualism, Wulff et al. discussed possible ways to classify defective grandfather clocks as an example.

Suppose that one examines how people who work in a repair shop might classify clocks. The receptionist, who knows very little about the workings of the clocks, might classify them descriptively. Thus, some clocks would be placed together because they do not work after being wound; others have broken faces, arms or other parts; and still others do not keep time accurately. Another person who might classify the grandfather clocks would be the bookkeeper of the shop. This individual might classify the clocks according to the manufacturer and cost of the clock. A third person who might classify the clocks is the repairman. He might organize clocks anatomically into those with accumulated dirt impeding their normal functioning, those needing replacement parts, and those with weighting mechanisms that have become unbalanced. Finally, the owner of the repair shop, when reporting back to various manufacturers about the causes of clock malfunctions, might classify the clocks etiologically. That is, she might report about clocks that have had little care, clocks that become worn over various time intervals of ownership, and clocks that developed problems after being moved or damaged.

Which of these classifications—descriptive, cost oriented, anatomical, or etiological—is the true or best classification of defective grandfather clocks? From the nominalist perspective, none of these classifications is inherently the best. Each of these classifications is imposed by the needs of the particular individual using the classification. Each classification serves a function. For any particular function, one classificatory system may be preferable. But none of these is the true classification of defective clocks.

Notice that this apocryphal classification of defective clocks is analogous to the approaches to defining disease suggested by Scadding: clinical-descriptive, morbid anatomical and etiological. The cost oriented classification was simply added as an analogy to how medical classifications are used by the insurance industry in the USA.

Wulff's defective clock analogy was borrowed from the British philosopher, John Locke. Locke had argued that classificatory systems are inherently nominalist, even though the ultimate goal is often essentialist:

> Therefore we in vain pretend to rank things in sorts, and dispose them into certain classes, under names, by their *real essences*, that are so far from our discovery or comprehension. (Wulff et al., 1986, p. 75)

In this regard, it is interesting to contrast Locke's view of classification to those of his physician friend, Thomas Sydenham. Believing in an essentialist view of disease, Sydenham made the following statement which has been quoted repeatedly since then:

> Nature, in the production of disease, is uniform and consistent ... The selfsame phenomena that you observe in the sickness of a Socrates you would observe in the sickness of a simpleton.

In other words, diseases do exist. They do have an essence. It is the business of medical research to discover what these essences are. It was the belief in this essentialist perspective that led nineteenth century researchers to solve the etiological issues associated with dementia paralytica.

Scadding and Wulff et al. warned about the dangers of essentialist thinking when applied to disease. For instance, Scadding noted that often we mistake the disease for the cause of the disease. Noguchi and Moore (Quetel, 1990) discovered the syphilitic bacilli in the brains of individuals with paresis. Hence, we might say that paresis occurs when syphilis invades the central nervous system. But the last sentence is misleading. Syphilis is not an organism. The bacteria, *Treponema pallidum*, is an organism and it could be said to invade the central nervous system. But even if this bacteria were present in the brain of an individual, that presence does not mean that the individual has paresis. To have paresis the individual must manifest the characteristic symptoms and anatomical changes associated with paresis.

Wulff et al. end their discussion of nominalism vs. essentialism with the following statement:

> The philosophical problem which underlies the discussion in this chapter is the age-old dispute about universals, and we have tried to navigate between the Scylla of essentialism (or Platonism) and the Charybdis of extreme nominalism. Essentialism underlines correctly that any classification of natural phenomena must reflect the realities of nature, but it ignores the fact that classifications also depend on our choice of criteria and that this choice reflects our practical interests and the extent of our knowledge. Nominalism, on the other hand, stresses correctly the human factor, but the extreme nominalist overlooks that classifications are not arbitrary but must be molded on reality as it is. (Wulff et al., 1986, pp. 87–88)

As mentioned earlier, defining the concepts of "disease" and/or "disorder" raise complicated issues, and the preceding discussion does not adequately cover the literature. One American author, in particular, has attracted substantial attention in the 1990s for his writings on this issue. Wakefield (1992, 1993) initially addressed this issue by providing a detailed critique of the definition of mental disorder that appeared in the *DSM-III*. Following this seminal paper, other theoretical papers (Wakefield, 1997a, 1997b) proposed a "harmful dysfunction" view of how to define mental disorders. A special section of the *Journal of Abnormal Psychology* has been devoted to a discussion of Wakefield's ideas .

Besides Wakefield's writings, there are other important discussions of this definitional issue including an overview by Reznek (1987), a book by Wing (1978), and an edited book on philosophical issues associated with classification (Sadler, Wiggins & Schwartz, 1994).

4.03.9.3 Two Views of a Hierarchical System of Diagnostic Concepts

Categories in the classification of mental disorders are organized hierarchically. This structural arrangement is commonly recognized but, since the publication of the *DSM-III*, two different views about this hierarchical structure have been discussed. Since these two views are often confused, the next section briefly discussed them.

The first approach to the meaning of hierarchy is nested set approach. Consider, for instance, the *DSM-II* classification of mental disorders. In this system, there are two broad categories of disorders: (I) psychotic disorders and (II) nonpsychotic disorders. The psychotic disorders are further subdivided into (I.A) the organic disorders and (I.B) the nonorganic disorders. The nonorganic psychotic disorders in the *DSM-II* were subdivided into three categories: (I.B.1) schizophrenic disorders, (I.B.2) major affective disorders, and (I.B.3) paranoid disorders. Then the schizophrenic disorders were subdivided into various subtypes: (I.B.1.a) simple type, (I.B.1.b) hebephrenic type, (I.B.1.c) catatonic type, and so on. Notice that this organization of mental disorders has a similar outline to the organization of categories in the biological classification. Any patient, for instance, who would be diagnosed as being hebephrenic, would also be considered as having schizophrenia disorder (the next highest, inclusive category) as well as having a psychotic disorder (an even higher level, inclusive category). Thus, this approach to hierarchy is called a nested set approach because the categories low in the system refer to sets of patients that are included (nested) in higher order categories. This approach parallels the

classification of biological organisms in which any animal who is a member of the species *Felis catus* (housecat) is also member of the genus *Felis* (cats) and a member of an even higher order category of *Mammalia* (mammal).

The other approach to hierarchy is called a pecking order view. This view can be best understood by making analogy to the hierarchical organization of rank among military officers. A colonel is higher in rank than a major who in turn is higher in rank than a lieutenant. In this pecking order structure, a colonel can give orders to a lieutenant, but a lieutenant cannot issue orders to a colonel. Thus, the pecking order in military rank concerns lines of authority. A colonel has authority over a lieutenant. Notice, however, that there is no membership nesting in these categories. If a particular individual is a lieutenant, then that individual is not a colonel even though a colonel is higher in the hierarchy than a lieutenant.

To understand how this analogy to the hierarchical arrangement of military rank can be applied to psychiatric classification, consider the following order of general mental disorders:

organic mental disorders
schizophrenic disorders
affective disorders
anxiety disorders
personality disorders.

In terms of the pecking order meaning of hierarchy, this order means that disorders higher in this order should be diagnosed over disorders lower in the hierarchy. Thus, in terms of standard diagnostic practice, the presence of organic mental disorders should be ruled out first, then the schizophrenic disorders, then affective disorders, etc. This principle of diagnostic precedence is analogous to the authority relationship among different levels of rank in the military.

Notice that the pecking order relationship among these five general mental disorders also carries another implication. If a patient has an organic mental disorder, the patient can (and often does) have the symptoms associated with disorders that are lower in the hierarchy. Thus, a patient with Alzheimer's disease can develop hallucinations like a schizophrenic, can have marked sleep disturbance like someone who is depressed, can develop a fear of leaving the house like someone with anxiety disorder, and show the rigidity and need to be controlled like someone with an obsessive-compulsive personality disorder. However, a patient with anxiety disorder such as agoraphobia should not show the disturbed memory patterns and disorientation of a patient with an organic mental disorder. The important point to note is that disorders placed higher in this pecking order view of hierarchy can explain any of the symptoms of disorders lower in the hierarchy; however, the reverse should not occur. There should be symptoms that will be manifest in patients with schizophrenia that will not occur in patients with personality disorders.

The pecking order approach to the hierarchical arrangement of mental disorder categories was popularized by Foulds and Bedford (1975). The specifics of their approach to the classification of mental disorders differs from that presented above, but the general outline is the same. An important corollary to this pecking order view of the hierarchical arrangement among mental disorder categories is that this view suggests that there will be a strong severity dimension in any descriptive approach to the classification of these disorders. Mental disorders higher in this system will be associated with many more symptoms than are mental disorders lower in this system. Descriptive studies of psychopathology have repeatedly found a strong severity dimension that the pecking order view would predict.

4.03.9.4 Problem of Diagnostic Overlap/ Comorbidity

When discussing the categorical model of classification, one of the tenets that was attributed to the model stated: "In the borderline areas where categories may overlap, the number of overlapping patients should be relatively small." Diagnostic overlap refers to the relative percentage of patients with one diagnosis who also meet the criteria for another diagnosis. As the tenet above states, some diagnostic overlap is expected. But the relative amount of overlap should be small.

One terminological note should be made before proceeding. The literature on this issue is grouped under the general heading of comorbidity. This term is from the medical literature, because it is well recognized that some medical disorders tend to go together. For instance, individuals who develop AIDS are relatively likely to develop yeast infections, sarcomas, and other disorders because of their compromised autoimmune system. The term comorbidity refers to the pattern of co-occurrences of these medical disorders. However, because the concept of comorbidity implies the acceptance of a disease model, the preferred term in this chapter will be "diagnostic overlap."

One of the earliest studies that focused on diagnostic overlap was by Barlow, DiNardo, Vermilyea, Vermilyea, and Blanchard (1986).

These investigators reported on 126 patients who were referred for the treatment of anxiety. These patients were administered structured interviews. Of the 126 patients interviewed, 108 were assigned one of seven diagnoses that fit within the anxiety/affective disorder spectrum. Of these 108 patients 65% were given at least one additional diagnosis. This is a high level of diagostic overlap and apparently was much higher than these researchers had expected *a priori*.

A large number of other empirical studies have confirmed the high levels of diagnostic overlap using the *DSM-III* and subsequent classifications. Many of these studies are discussed in an excellent review by Clark, Watson, and Reynolds (1995). Examples of the results found in their review are listed below. For example, these reviewers noted one study of personality disorder diagnoses in a state hospital population that found that these patients met the diagnostic criteria for an average of 3.75 Axis II disorders. In addition to the personality disorders, the depressive disorders also have striking overlap with many other disorders. For instance, over half of the patients with major depressive disorder as well as patients with dysthymic disorder were found to have at least one co-occuring mental disorder. Depression shows significant overlap even with disorders that one might not expect overlap. For instance, a sample of antisocial personality disorder patients showed that one-third of these individuals also had a depressive diagnosis (Dinwiddie & Reich, 1993). These antisocial patients, less surprisingly, also had high rates of alcoholism (76%) and other substance use (63%). Even broad epidemiological studies on normal community samples show high rates of diagnostic overlap. Clark et al. reported that, in two national studies, over half of the individuals who had one mental disorder diagnosis had at least one more diagnosis. In some samples, the rate of diagnostic overlap is even higher. For instance, a study of suicidal patients showed that these individuals averaged about four mental disorders (Rudd, Dahm, & Rajab, 1993).

Together, these and many other studies suggest that the number of overlapping patients among mental disorder diagnoses is not small at all. Instead, diagnostic overlap is a standardly occurring phenomenon. Blashfield, McElroy, Pfohl, and Blum (1994) studied 151 patients who had been administered a structured interview to assess personality disorders. In this sample, only 24% of the patients met the criteria for one and only one personality disorder. Exactly the same percentage met the diagnostic criteria for at least four personality disorders, with 6% meeting the criteria for six of the eleven *DSM-III-R* personality disorders! When Blashfield et al. attempted to identify prototypic patients (i.e., individuals who met at least eight of the diagnostic criteria for a specific disorder), they found that only 15% of the patients would qualify as prototypic. However, most of these individuals also satisfied the diagnostic criteria for other disorders. When a prototypic patient was defined as an individual with eight or more criteria for a personality disorder and the lack of an additional personality diagnosis, only 1% of the patients were prototypes. In effect, patients with mixed symptom patterns are much more typical than are patients who represent relatively pure forms of a disorder.

Clark, Watson and Reynolds (1995) suggested that there are three related issues associated with the problem of diagnostic overlap. The first issue concerns the hierarchical organization of categories. As discussed earlier, one view of hierarchy is a pecking order approach. This view was implicitly adopted by the *DSM-III* because a number of exclusion rules were included in the diagnostic criteria for different disorders, so that a lower diagnosis would not be made if a higher order diagnosis was possible. However, research on the exclusionary criteria suggested that these criteria were arbitrary. Exclusionary criteria were mostly deleted in the *DSM-III-R* and the *DSM-IV*.

The second issue associated with the diagnostic overlap issue concerns the heterogeneity with diagnostic categories. In effect, the extensive diagnostic overlap suggests that the definitions of the various mental disorders are too broad and inclusive. Evidence of excessive heterogeneity comes from other sources, according to Clark et al. For instance, direct studies of the variability in symptom patterns have shown high levels of variability within disorders such as schizophrenia, depression, eating disorders and anxiety disorders. Another line of evidence of heterogeneity is the frequency with which mixed or atypical diagnoses are used. For instance, Mezzich, Fabrega, Coffman, and Haley (1989) found that the majority of patients with a dissociative disorder fit the criteria for an atypical dissociative disorder.

The third issue associated with the comorbidity finding is the increasing support associated with replacing the categorical approach to classification with a dimensional view. In the discussion of these two models, it was noted that the dimensional model is the simpler of the two. Unless there is clear evidence of the existence of discrete categories, a dimensional approach is the more parsimonious.

A number of researchers, when confronted with high rates of diagnostic overlap, have suggested that a dimensional model should be used. For instance, in the area of the personality disorders where some of the highest rates of diagnostic overlap have been found, interest in a "Big Five" dimensional approach to the personality disorders has been attracting increasing support. Another example in which a dimensional model is gaining popularity is in the subclassification of schizophrenia. The classic Kraepelinian subtypes do not generate sufficiently homogeneous categories. Various dimensional schemes (process–reactive, paranoid–nonparanoid, positive vs. negative symptoms) have more empirical support than the use of categorical subtypes.

4.03.10 CONCLUDING COMMENTS

This chapter provides a rather simplified overview of the issues associated with the classification of psychopathology. It attempts to help the reader gain a better understanding of classification by starting with a reasonably detailed history of classificatory systems, and thereby give some idea how many of the features of contemporary classificatory systems have evolved. The text also presents a succinct and readable presentation of four issues that currently attract a reasonable degree of attention in the literature. However, justice has not been done to many of the other complex issues that face both clinicians and scientists interested in classification, such as whether to use semistructured interviews as the "gold standard" for measurement (Mezzich, 1984), the role of values in the diagnostic practice of mental health professionals, whether or not certain mental disorders are sexually or racially biased (Busfield, 1996; Nuckolls, 1992; Widom, 1984), the relevance of life-span measures of psychopathology (Roff & Ricks, 1970), and the problem of focusing on the individual patient as the basic unit of psychopathology (as opposed to families or systems or interpersonal relationship patterns) (Clarkin & Miklowitz, 1997; Francis, Clarkin & Ross, 1997; Williams, 1997)

Like any general topic, the classification of psychopathology becomes a very complex topic when analyzed in detail.

> Perhaps I believe that the world can get forward most by clearer and clearer definitions of fundamentals. Accordingly I propose to stick to the tasks of nomenclature and terminology, unpopular and ridicule-provoking though they may be. (Southard, as quoted by Menninger, 1963, p. 3)

4.03.11 REFERENCES

American Psychiatric Association (1933). Notes and comments: Revised classified nomenclature of mental disorders. *American Journal of Psychiatry, 90,* 1369–1376.

American Psychiatric Association (1952). *Diagnostic and statistical manual of mental disorders* (1st ed.). Washington, DC: Author.

American Psychiatric Association (1968). *Diagnostic and statistical manual of mental disorders* (2nd ed.). Washington, DC: Author.

American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed. Rev.). Washington, DC: American Psychiatric Press.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Press.

Alexander, F. G., & Selesnick, S. T. (1966). *The history of psychiatry.* New York: Harper and Row.

Austin, T. J. (1859/1976). *A practical account of general paralysis.* New York: Arno Press.

Barlow, D. H., DiNardo, P. A., Vermilyea, B. B., Vermilyea, J., & Blanchard, E. B. (1986). Co-morbidity and depression among the anxiety disorders: Issues in diagnosis and classification. *Journal of Nervous and Mental Disease, 174,* 63–72.

Barsalou, L. W. (1992). *Cognitive psychology: An overview for cognitive scientists.* Hillsdale, NJ: Erlbaum.

Bayer, R. (1981). *Homosexuality and American psychiatry.* New York: Basic Books.

Berrios, G. E. (1993). Personality disorders: A conceptual history. In P. Tyrer & G. Stein (Eds.), *Personality disorder reviewed* (pp. 17–41). London: Gaskell.

Berrios, G. E., & Hauser, R. (1988). The early development of Kraepelin's ideas on classification: A conceptual history. *Psychological Medicine, 18,* 813–821.

Blashfield, R. K. (1991). Models of psychiatric classification. In M. Hersen & S. M. Turner (Eds.), *Adult psychopathology and diagnosis* (2nd ed. pp. 3–22). New York: Wiley.

Blashfield, R. K., McElroy, R. A., Pfohl, B., & Blum, N. (1994). Comorbidity and the prototype model. *Clinical Psychology: Science and Practice, 1,* 96–99.

Blustein, B. E. (1991). *Preserve your love for science: Life of William A. Hammond, American neurologist.* Cambridge, UK: Cambridge University Press.

Braslow, J. T. (1995). Effect of therapeutic innovation on perception of disease and the doctor–patient relationship: A history of general paralysis of the insane and malaria fever therapy, 1910–1950. *American Journal of Psychiatry, 152,* 660–665.

Busfield, J. (1996). *Men, women and madness: Understanding gender and mental disorder.* New York: New York University Press.

Campbell, E. J. M., Scadding, J. G., & Roberts, R. J. (1979). The concept of disease. *British Medical Journal, 2,* 757–762.

Cantor, N., Smith, E. E., French, R. D., & Mezzich, J. (1980). Psychiatric diagnosis as a prototype categorization. *Journal of Abnormal Psychology, 89,* 181–193.

Clark, L. A., Watson, D., & Reynolds, S. (1995). Diagnosis and classification of psychopathology: Challenges to the current system and future directions. *Annual Review of Psychology, 46,* 121–153.

Clarkin, J. F., & Mikowitz, D. J. (1997). Marital and family communication difficulties. In T. A. Widiger, A. J. Francis, H. A. Pincus, R. Ross, M. B. First, & W. Davis (Eds.), *DSM-IV sourcebook* (Vol. 3, pp. 631–672). Washington, DC: American Psychiatric Press.

Dinwiddie, S. H., & Reich, T. (1993). Attribution of antisocial symptoms in coexistent antisocial personality disorder and substance abuse. *Comprehensive Psychiatry, 34,* 235–242.

Essen-Moller, E. (1971). Suggestions for further improvement of the international classification of mental disorders. *Psychological Medicine, 1,* 308–311.

Everitt, B. S. (1974). *Cluster analysis.* New York: Halstead Press.

Feighner, J. P., Robins, E., Guze, S., Woodruff, R. A., Winokur, G., & Munoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry, 143,* 57–63.

Feinstein, A. R. (1967). *Clinical judgment.* Huntington, VA: Krieger.

Foulds, G. A., & Bedford A. (1975). Hierarchy of classes of personal illness. *Psychological Medicine, 5,* 181–192.

Francis, A. J., Clarkin, J. F., & Ross, R. (1997). Family/relational problems. In T. A. Widiger, A. J. Francis, H. A. Pincus, R. Ross, M. B. First, & W. Davis (Eds.), *DSM-IV sourcebook* (Vol. 3, pp. 521–530). Washington, DC: American Psychiatric Press.

Goffman, E. (1961). *Asylums.* London: Penguin.

Grinker, R. R., Werble, B., & Drye, R. C. (1968). *The borderline syndrome.* New York: Basic Books.

Hammond, W. A. (1883). *A treatise on insanity in its medical relations.* New York: Appleton.

Hempel, C. G. (1965). *Aspects of scientific explanation.* New York: Free Press.

Hull, D. L. (1988). *Science as a process.* Chicago: University of Chicago Press.

Jardine, N., & Sibson, R. (1971). *Mathematical taxonomy.* New York: Wiley.

Kendell, R. E. (1975). *The role of diagnosis in psychiatry.* Oxford, UK: Blackwell.

Kendell, R. E. (1976). The concept of disease. *British Journal of Psychiatry, 128,* 508–509.

Kendell, R. E. (1979). Alcoholism: A medical or a political problem. *British Medical Journal, 1,* 367–381.

Kendell, R. E. (1986). What are mental disorders? In A. M. Freedman, R. Brotman, I. Silverman, & D. Huston (Eds.), *Issues in psychiatric classification* (pp. 23–45). New York: Human Sciences Press.

Kendell, R. E. (1991). Relationship between the DSM-IV and ICD-10. *Journal of Abnormal Psychology, 100,* 297–301.

Kirk, S. A., & Kutchins, H. (1992). *The selling of DSM: The rhetoric of science in psychiatry.* Hawthorne, NY: Walter deGruyter.

Klerman, G. L. (1978). The evolution of a scientific nosology. In J. C. Shershow (Ed.), *Schizophrenia: Science and practice* (pp. 99–121). Cambridge, MA: Harvard University Press.

Kraepelin, E. (1902/1896). *Clinical psychiatry: A text-book for students and physicians* (6th ed., translated by A. R. Diefendorf). London: Macmillan.

Kreitman, N. (1961). The reliability of psychiatric diagnosis. *Journal of Mental Science, 107,* 878–886.

Lacan, J. (1977). *Ecruits: A selection.* New York: Norton.

Laing, R. D. (1967). *The politics of experience.* London: Penguin.

Lorr, M. (1966). *Explorations in typing psychotics.* New York: Pergamon.

Matza, D. (1969). *Becoming deviant.* Englewood Cliffs, NJ: Prentice-Hall.

Meehl, P. E. (1995). Bootstraps taxometrics: Solving the classification problem in psychopathology. *American Psychologist, 50,* 266–275.

Menninger, K. (1963). *The vital balance.* New York: Viking.

Mezzich, J. E. (1984). Diagnosis and classification. In S. M. Turner & M. Hersen (Eds.), *Adult psychopathology and diagnosis* (pp. 3–36). New York: Wiley.

Mezzich, J. E., Fabrega, H., Coffman, G. A., & Haley, R. (1989). DSM-III disorders in a large sample of psychiatric patients: Frequency and specificity of diagnosis. *American Journal of Psychiatry, 146,* 212–219.

Mezzich, J. E., Fabrega, H., Mezzich, A. C., & Coffman, G. A. (1985). International experience with DSM-III. *Journal of Nervous and Mental Disease, 173,* 738–741

Nelson, G., & Platnick, N. (1981). *Systematics and biogeography. Cladistics and vicariance.* New York: Columbia University Press.

Nuckolls, C. W. (1992). Toward a cultural history of the personality disorders. *Social Science and Medicine, 35,* 37–47.

Plaut, F. (1911). *The Wasserman sero-diagnosis of syphilis in its application to psychiatry.* New York: Journal of Nervous and Mental Disease Publishing Company.

Project Match Research Group (1997). Matching alcoholism treatments to client heterogeneity: Project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol, 58,* 7–29.

Quetel, C. (1990). *History of syphilis.* Baltimore: Johns Hopkins University Press.

Raines, G. N. (1952). Forward. In American Psychiatric Association, *Diagnostic and statistical manual of mental disorders* (1st ed., pp. v–xi). Washington, DC. American Psychiatric Assocation.

Regier, D. A., Kaelber, C. T., Roper, M. T., Rae, D. S., & Sartorius, N. (1994). The ICD-10 clinical field trial for mental and behavioral disorders: Results in Canada and the United States. *American Journal of Psychiatry, 151,* 1340–1350.

Reznek, L. (1987). *The nature of disease.* London: Routledge & Kegan Paul.

Roff, M., & Ricks, D. F. (Eds.) (1970). *Life history research in psychopathology.* Minneapolis, MN: University of Minnesota Press.

Rosenhan, D. L. (1973). On being sane in insane places. *Science, 179,* 250–258.

Roth, A., & Fonagy, P. (1996). *What works for whom: A critical review of psychotherapy research.* New York: Guilford.

Rudd, M. D., Dahm, P. F., & Rajab, M. H. (1993). Diagnostic comorbidity in persons with suicidal ideation and behavior. *American Journal of Psychiatry, 147,* 1025–1028.

Russell, J. A. (1991). In defense of a prototype approach to emotion concepts. *Journal of Personality and Social Psychology, 60,* 37–47.

Russell, J. A., & Fehr, B. (1994). Fuzzy concepts in a fuzzy hierarchy: Varieties of anger. *Journal of Personality and Social Psychology, 67,* 186–205.

Rutter, M., Shaffer, D., & Shepard, M. (1975). *A multiaxial classification of child psychiatry disorders.* Geneva, Switzerland: World Health Organization.

Sadler, J. Z., Wiggins, O. P., & Schwartz, M. A. (1994). *Philosophical perspectives on psychiatric diagnostic classification.* Baltimore: Johns Hopkins University Press.

Sarbin, T. R. (1997). On the futility of psychiatric diagnostic manuals (DSMs) and the return of personal agency. *Applied and Preventive Psychology, 6,* 233–243.

Scadding, J. G. (1959). Principles of definition in medicine with special reference to chronic bronchitis and emphysema. *Lancet, 1,* 323–325.

Scadding, J. G. (1969). Diagnosis: The clinician and the computer. *Lancet, 2,* 877–882.

Skinner, H. A. (1986). Construct validation approach to psychiatric classification. In T. Millon & G. L. Klerman (Eds.), *Contemporary directions in psychopathology: Toward the DSM-IV* (pp. 307–331). New York: Guilford Press.

Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical taxonomy.* San Francisco: Freeman.

Spitzer, R. L., Endicott, J., & Robins, E. (1975). Research

diagnostic criteria. *Archives of General Psychiatry, 35*, 773–782.
Spitzer, R. L., & Fleiss, J. L. (1974). A re-analysis of the reliability of psychiatric diagnosis. *British Journal of Psychiatry, 125*, 341–347.
Spitzer, R. L., & Williams, J. B. W. (1987). Revising DSM-III: The process and major issues. In G. L. Tischler (Ed.), *Diagnosis and classification in psychiatry* (pp. 425–434). New York: Cambridge University Press.
Spitzka, E. C. (1883). *Insanity: Its classification, diagnosis and treatment*. New York: Bermingham.
Stangl, D., Pfohl, B., Zimmerman, M., Bowers, W., & Corenthal, C. (1985). Structured interview for the DSM-III personality disorders. *Archives of General Psychiatry, 42*, 519–596.
Stengel, E. (1959). Classification of mental disorders. *Bulletin of the World Health Organization, 21*, 601–663.
Stromgren, E. (1991). A European perspective on the conceptual approaches to psychopathology. In A. Kerr & H. McClelland (Eds.), *Concepts of mental disorders: A continuing debate* (pp. 84–90). London: Gaskell.
Szasz, T. (1961). *The myth of mental illness*. New York: Hoeber-Harper.
Veith, I. (1965). *Hysteria: The history of a disease*. Chicago: University of Chicago Press.
Wakefield, J. C. (1992). The concept of mental disorder: On the boundary between biological facts and social values. *American Psychologist, 47*, 373–388.
Wakefield, J. C. (1993). Limits of operationalization: A critique of Spitzer and Endicott's (1978) proposed operational criteria for mental disorder. *Journal of Abnormal Psychology, 102*, 160–172.
Wakefield, J. C. (1997a). Diagnosing DSM-IV—Part I: DSM-IV and the concept of disorder. *Behavioral Research and Therapy, 35*, 633–649.
Wakefield, J. C. (1997b). Diagnosing DSM-IV—Part II: Eysenck (1986) and the essentialist fallacy. *Behavioral Research and Therapy, 35*, 651–665.
Walker, L. (1987). Inadequacies of the masochistic personality disorder diagnosis for women. *Journal of Personality Disorders, 1*, 183–189.
Widom, C. (Ed.) (1984). *Sex, roles and psychopathology*. New York: Plenum.
Williams, J. B. W. (1997). The DSM-IV multiaxial system. In T. A. Widiger, A. J. Francis, H. A. Pincus, R. Ross, M. B. First, & W. Davis (Eds.), *DSM-IV sourcebook* (Vol. 3). Washington, DC: American Psychiatric Press.
Wing, J. K. (1978). *Reasoning about madness*. Oxford, UK: Oxford University Press.
Woodruff, R. A., Goodwin, D. W., & Guze, S. B. (1974). *Psychiatric diagnosis*. New York: Oxford University Press.
World Health Organization (1948). *Manual of the international statistical classification of diseases, injuries, and causes of death*. Geneva, Switzerland: Author.
World Health Organization (1957). Introduction to *Manual of the international statistical classification of diseases, injuries, and causes of death* (7th ed.). Geneva, Switzerland: Author.
World Health Organization (1978). *Mental disorders: Glossary and guide to their classification in accordance with the ninth revision to the International Classification of Diseases*. Geneva, Switzerland: Author.
World Health Organization (1992). *The ICD-10 classification of mental and behavioral disorders: Clinical descriptions and diagnostic guidelines*. Geneva, Switzerland: Author.
World Health Organization (1993). *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research*. Geneva, Switzerland: Author.
Wulff, H. R., Pedersen, S. A., & Rosenberg, R. (1986). *Philosophy of medicine*. Boston: Blackwell Scientific.
Wyer, R. S., & Srull, T. K. (1989). *Memory and cognition in a social context*. Hillsdale, NJ: Erlbaum.
Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*, 338–353.
Zubin, J. (1967). Classification of behavior disorders. *Annual Review of Psychology, 28*, 373–406.

4.04
Clinical Interviewing

EDWARD L. COYLE
Oklahoma State Department of Health, Oklahoma City, OK, USA
and
DIANE J. WILLIS, WILLIAM R. LEBER, and JAN L. CULBERTSON
University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA

4.04.1 PURPOSE OF THE CLINICAL INTERVIEW 82
4.04.1.1 Gathering Information for Assessment and Treatment 82
4.04.1.2 Establishing Rapport for Assessment and Treatment 82
4.04.1.3 Interpersonal Style/Skills of the Interviewer 82
4.04.1.4 Structuring the Interview 83
4.04.1.4.1 Setting variables 83
4.04.1.4.2 Preparing for the patient 84
4.04.1.5 Introductory Remarks 84
4.04.1.6 How to Open the Interview 85
4.04.1.7 The Central Portion of the Interview 85
4.04.1.8 Closing the Interview 86
4.04.1.9 The Collateral Interview 86
4.04.2 DEVELOPMENTAL CONSIDERATIONS IN INTERVIEWING 87
4.04.2.1 Interviewing Children (Preschool Age through Older Elementary) 87
4.04.2.2 Interviewing Parents 87
4.04.2.3 Social Context 88
4.04.2.4 Developmental Context 89
4.04.2.5 Direct Interview of Children 89
4.04.2.6 Adolescents 90
4.04.2.6.1 Separation–individuation 91
4.04.2.6.2 Resolving conflict with authority figures 91
4.04.2.6.3 Peer group identification 91
4.04.2.6.4 Realistic appraisal and evaluation of self-qualities 92
4.04.2.7 Interviewing Young Adults (18–40 Years) 92
4.04.2.8 Interviewing Adults in Middle Adulthood (40–60 Years) 93
4.04.2.9 Interviewing Older Adults (60–70 Years) 93
4.04.2.10 Interviewing In Late Adulthood (70 Years–End of Life) 93
4.04.3 INTERVIEWING SPECIAL POPULATIONS OR SPECIFIC DISORDERS 94
4.04.3.1 Interviewing Depressed Patients 94
4.04.3.2 Interviewing Anxious Patients 94
4.04.4 SUMMARY 95
4.04.5 REFERENCES 96

CHAPTER 16

OBJECTIVE PERSONALITY ASSESSMENT

Laura S. Keller
James N. Butcher
Wendy S. Slutske

In his analysis of the impact of technological change on individuals and society, Mesthene (1970) described two extreme views of technology: as an "unalloyed blessing for man and society" or alternatively as "an almost unmitigated curse" (pp. 16–17). Those who favor the latter view stress the impersonal and potentially dehumanizing aspects of technology. Proponents of technological change emphasize its promise for "freeing man to pursue creative productivity" (Williams, 1977, p. 108). As an important component of psychology's technology, objective personality assessment has been both criticized and praised on the basis of extreme value positions such as these. Realistically, however, technological developments should be evaluated only in relation to their relevance and utility for the achievement of valued goals. Objective methods of personality assessment and their logical outgrowth, automated assessment methods, are particularly well suited to the achievement of practical goals in both research and applied psychological settings.

Objective assessment devices are characterized by the structured and standardized nature of their test stimuli, response options available to the subject, and scoring and interpretive strategies. Objective tests can be reliably and accurately administered, scored, and in some cases actually interpreted by nonprofessional workers or even by a machine. The latter is possible because of the huge amount of research data available on many popular objective personality instruments. These data have been incorporated into standardized systems of interpretive rules relating test scores to meaningful extra-test criteria. The reliability and standardization of data collected through objective personality inventories are ideal for research applications. In addition, the efficiency of such assessment methods is increasingly important in applied settings; limited funding and the current emphasis on accountability are making the cost/benefit ratios of various psychological technologies a standard concern.

HISTORICAL FOUNDATIONS

Throughout the history of mankind, various informal or formal methods have been used to classify, select, or predict the behavior of other human beings. As Ben-Porath and Butcher (in press) describe in their historical review, descriptions of personality assessment techniques can be found in the Bible, in ancient Chinese selection procedures, and in early Roman and Greek writings. They note three major divisions of personality assessment techniques, categorized by the type and source of data: (a) behavioral observations, (b) somatic examinations, and (c) verbal examinations.

Behavioral and Somatic Examinations

Behavioral observations lie at the base of most of our informal, personal strategies for assessing the character of those around us. They have also been

used extensively in formal strategies for describing personality or predicting future behavior. Galton (1884) wrote of assessing individual character through observation of behavior, and Hartshorne and May (1928) described a series of behavioral tests designed to measure the personality trait of "honesty." The low correlations between the various single-situation honesty measures resulting from the latter test serves to point out a basic principle in objective personality testing: single items, whether they be behavioral observations or verbal responses, tend to be very unreliable and poor predictors of other individual descriptors or future behavior. Thus, *aggregation of information,* such as by grouping several observations across several situations, grouping information obtained by several different assessment methods, or simply by adding individual items together to form scales, is basic to the development of reliable, useful, descriptive and predictive personality measures.

Behavioral observations, such as performance-based tests for job selection or teachers' ratings of child classroom behavior, are widely used today. However, the trend has been toward directly measuring and changing behaviors in question rather than inferring underlying personality dispositions or predicting nonsampled behaviors from such measures. Although the boundaries are fuzzy, the contemporary uses of behavioral observation techniques are probably covered more appropriately in other chapters of this handbook, so will not be explored here.

Similarly, the use of somatic examinations to assess personality has lost popularity in the past few decades. The heyday of somatic assessment in the 1800s revolved around the techniques of phrenology, a systematized method of describing individual character and mental faculties through examining the size, shape, and form of the skull (Davies, 1955). Both this invalidated technique and the "somatotype" personality descriptors proposed by Sheldon (1940) have fallen into disfavor, leaving the somatic approach to personality assessment largely unused today. As Ben-Porath and Butcher (in press) point out, this area may see some new growth in the future with the expansion of new psychophysiological measures.

Verbal Personality Tests

The third category of personality tests, verbal examinations, has dominated the field of personality assessment this century. Verbal tests are traditionally divided into *projective techniques,* distinguished primarily by the unstructured nature of their response options, and *objective techniques,* which provide the responder with a limited set of structured response options. Although we will not discuss projective techniques in this chapter, it should be noted that the distinction is not as clear as it initially sounds: even in "objective" tests, "there is a 'projective' element involved in interpreting and responding to these verbal stimuli which must be recognized, in spite of the fact that the test situation is very rigidly structured as regards the ultimate response possibilities permitted" (Meehl, 1945, p. 299).

Although contemporary objective techniques had their roots in rating scales completed by individuals other than the subject being described, questionnaires completed by the test subject soon overshadowed rating scales in popularity. These *self-report inventories* were aggregates of "conventional culturally crystallized questions to which the subject must respond in one of a few fixed ways" (Meehl, 1945, p. 296). The first self-report inventories were developed in applied settings to aid psychologists in practical decision-making tasks. The Woodworth Personal Data Sheet (PDS) was constructed during World War I for mass psychiatric screening of military draftees (Woodworth, 1920). Designed to function as a standardized clinical interview, the PDS was a paper-and-pencil test consisting of questions about neurotic symptoms. The subject's score on this test was the total number of symptoms he or she endorsed.

After the war, many other inventories were developed for a variety of predictive purposes and by a variety of methods. Many, like the PDS, were developed through *rational* methods: items were selected, written, and/or grouped according to the test developer's theory of which items and scales should work for his or her purpose. However, researchers began to find that a priori predictions of how particular items would relate to criteria of interest were often disconfirmed in practice: "Surprising as it may be, there is simply too little empirical support for the notion that a sophisticated theoretical background boosts scale validity" (Burisch, 1984, p. 215).

A different method of test construction, the *empirical criterion keying* approach, dealt with this problem by asserting that a person's response "constitutes an intrinsically interesting and significant bit of verbal behavior, the non-test correlates of which must be discovered by empirical means" (Meehl, 1945, p. 297). Rather than grouping items into scales according to preconceived theoretical notions of relevant content, the empirical test constructor uses the *method of contrasted groups* to select items according to their significance in distinguishing between groups of people differing on the criterion of interest. For example, in constructing a "depression" scale for the MMPI, the responses of a group of diagnosed depressives were contrasted with those of a group of "normals." Items for which the frequency of endorsement differed significantly between the two groups were included in the scale. The Minnesota Multiphasic Personality Inventory (MMPI) (Dahlstrom, Welsh, & Dahlstrom, 1972, 1975; Hathaway & McKinley, 1940) and its "normal" personality counterpart, the California Psychological Inventory (CPI) (Baucom, 1985; Eysenck, 1985; Gough, 1975; Megargee, 1972), were both constructed largely through methods of empirical criterion keying. Empirical or "external" methods of test development and validation are particularly popular for practical uses involving selection or classification of individuals into predefined groups (Burisch, 1984).

A third method of personality test construction has gained in popularity as the computer technology necessary for its execution has become available. In the *factor analytic, internal consistency,* or *inductive* approach to test construction, homogeneous scales are developed by factor analyzing the inter-item correlations among a large pool of personality items. This approach tends to be used more often by personality researchers hoping to uncover the "basic dimensions" of personality than by practitioners, at least initially. The Guilford-Zimmerman Temperament Survey (Guilford & Zimmerman, 1956) and the Comrey Personality Scales (Comrey, 1970) were developed through internal consistency methods. However, the factor analytic method is best exemplified in the work of Cattell, the developer of the Sixteen Personality Factor Questionnaire or 16PF (Cattell, Eber, & Tatsuoka, 1970).

Cattell attempted to provide a comprehensive description of personality by factor analyzing ratings of individuals on a set of trait items derived from the set of all personality trait names in the English dictionary and the psychiatric literature. Cattell described the resultant factors as "primary source traits" of personality, which Anastasi (1988, p. 542) noted is "a designation that seems to imply more universality and stability of results than appear justified by the antecedent research." However, the 16PF has proved to be a popular instrument for "normal population" personality assessment, and over the years a large body of research data has accrued on its use (Butcher, 1985; Zuckerman, 1985). More recently, the 16PF has been expanded to include dimensions of pathological personality functioning as well. This new instrument, the Clinical Analysis Questionnaire (CAQ), is designed to provide a comprehensive assessment of every meaningful dimension of both normal and pathological functioning (Krug, 1980). Several other adaptations of the 16PF have been developed for particular assessment purposes such as marriage counseling and job evaluation. However, the empirical validation research on these factorially developed instruments remains sketchy (Anastasi, 1988; Bolton, 1985; Butcher, 1985; Guthrie, 1985; McNair, 1978).

In the past two decades, test construction efforts have tended to combine several strategies in a multistep developmental process. Jackson (1971) presented a model test-construction strategy involving rational-theoretical selection of items based on the needs theory of Henry Murray (1938), followed by scale refinement through factorial internal consistency checks, and finally empirical validation against external criteria. Jackson's Differential Personality Inventory (DPI) (Jackson, 1972) and Personality Research Form (PRF) (Jackson, 1984) were constructed through this combination of rational, internal consistency, and empirical validation methodology. The inventories consist of relatively independent scales with high content homogeneity, and have also been designed to minimize the distorting effects of certain subject response styles. On reviewing the research literature of the past 10 years, however, we would tend to agree with Anastasi (1988) that "on the whole, the PRF is an excellent research instrument, but more information is needed to determine its effectiveness in practical situations" (p. 548).

Another of the "contenders to replace the MMPI monopoly" (Kleinmuntz, 1982, p. 242) is Millon's Clinical Multiaxial Inventory (MCMI). Like the PRF, the MCMI was developed using a mixture of scale construction methods. Item and scale content are linked to Millon's own clinical personality theory, but selection of the final item pool and scale composition were refined through a combination of internal consistency methods and external criterion checking. Millon set out to "draw upon the best featuresof the MMPI, minimize its limitations, and move forward to develop instruments that reflect advances of the past quarter of a century in psychopathology, diagnostic assessment, and test construction" (Millon, 1982, p. 1). Accordingly, the MCMI is distinguished by its brevity compared to comparable instruments, its link to a systematic clinical theory, the coordination of its interpretations with the current official psychiatric diagnostic system, its attempt to separate enduring personality characteristics from acute clinical symptom pictures, and its purported ability to discriminate among psychiatric groups. Because it was constructed with the latter goal in mind, the instrument is not recommended for use in "normal" populations or "for purposes other than diagnostic screening or clinical assessment" (p. 3). Although many authors (Butcher & Keller, 1984; Greer, 1984; Hess, 1985; Widiger,

1985; Widiger & Sanderson, 1987; Widiger, Williams, Spitzer, & Frances, 1985) have noted the lack of empirical validity information on the MCMI, in the past few years the data base on this instrument has been growing. While some researchers have found the MCMI to be useful and accurate for their purposes, broad concerns about its tendency to overpathologize compared to clinicians' judgments and its lack of correspondence to DSM-III diagnoses (American Psychiatric Association, 1980) have been raised. A revised form of the MCMI has been developed by Millon to incorporate new advances in his personality theory, to correspond to the revised diagnostic categories in the DSM-III-R, and to incorporate MCMI validation research (Millon, 1985a, 1986). The MCMI-II (Millon, 1987) includes two new personality disorder scales, new validity corrections, and a new item weighting system.

Despite the differences noted in test development strategies, in practice today most tests are either developed via a "mixed" strategy or, as is the case with the MMPI, their use has evolved to include a variety of approaches beyond their original developers' aims and methods. In general, research has failed to show that any particular one of these test construction methods results in significantly more valid or effective measurement devices than any other (Burisch, 1984; Goldberg, 1972; Hase & Goldberg, 1967).

CONTEMPORARY STATUS OF OBJECTIVE PERSONALITY TESTS

The proliferation of objective personality tests in the last half century attests to their practical advantages. For example, the ninth edition of the *Mental Measurements Yearbook* (Mitchell, 1985) categorizes 350 tests as personality instruments, 24.8% of all tests reviewed. Seventy-two percent of these were new since the last publication of the yearbook! Given these overwhelming numbers, we will not attempt in the space of this chapter to provide an overview of available tests. There are many other sources for such reviews including, of course, the *Ninth Mental Measurements Yearbook*. Anastasi (1988) also provides brief descriptions of many personality tests and measurement issues in her textbook on psychological testing. As we described above, a wide variety of methods and measures has been subsumed under the broad and often vague category of "personality tests." Many of these methods, such as projective techniques, interest inventories, behavioral observations, diagnostic interviews, and peer assessments, are covered elsewhere in this volume. At present, self-report inventories represent by far the largest category of objective personality tests. Although it is not possible for us to provide a detailed evaluation of recent research and clinical applications of objective personality tests, we will try to provide a brief overview of general areas of current use. Our survey is designed to update the review of objective personality research provided by Butcher and Owen in 1978. Our survey of literature included all articles using one or more of the MMPI, MCMI, DPI, PRF, 16PF, or CAQ found in a computer search (MEDLINE and PSYCH ABSTRACTS) of the published literature from 1978 to early 1988. We recognize that some publications may have been missed inadvertently, and that the delay between publication and listing in these data bases has forced us to exclude much of the literature published in the past year. We chose these particular instruments because they are the most widely used objective measures, they are designed to cover a wide variety of personality dimensions, they represent different major methods of test development, and they include the measures used by Butcher and Owen (1978) in their review.

Survey of Recent Literature

Even a cursory look at the current literature indicates that either omnibus personality inventories or one or more of their scales are used as criterion measures for clinical and normal sample research on a routine basis. The MMPI scales, in particular, are often used as the standard against which other devices are compared, and current psychotherapeutic treatment outcome studies almost invariably use some aspect of the MMPI as part of their assessment strategy. Changes in MMPI or MCMI indices from pretreatment to posttreatment are often used as measures of treatment efficacy, and many studies utilize pretreatment MMPI or other test administrations to investigate the relationship between personality factors or level/type of pathology and subsequent treatment outcome.

In their comprehensive review of objective personality inventory research, Butcher and Owen (1978) noted several major areas of current interest: detection of alcohol and drug problems, personality assessment of medical patients, cross-cultural and subcultural applications, and development of various short forms and alternative forms of the MMPI. These content domains have continued to attract research interest. In particular, alcohol and drug abuse have been heavily researched. Much of the work continues to center on

validation of the MMPI MacAndrew Addiction scale across diverse populations, although many other scales and strategies for differentiating alcoholics and/or drug addicts from various other groups are reported and sometimes compared to the MAC scale. The search continues for useful MMPI typologies to identify subgroups of chemical abusers sharing common treatment-relevant characteristics, and complex multivariate methods such as cluster analysis are increasingly popular tools for researchers searching for patient *types* rather than *the* addictive personality.

The current research on alcoholism is illustrative of a trend that can be seen across all areas of personality assessment research, namely the growing concern with demographic/cultural differences that might modify the validity of objective descriptors and predictors. Many of the research reports we saw were investigations of the generalizability of various scales or typologies to different socioeconomic classes, ethnic groups, and other nationalities, as well as of their generalizability across age and sex. International applications of the MMPI, for example, have dramatically increased in the past 10 years. There have been over 115 foreign language translations of the MMPI. The test is presently being used for clinical and research purposes in many countries, including such unlikely nations as the Soviet Union (Valsiner, 1982) and China (Song, 1981) that in the past had not been open to using personality scales for individual personality assessment. Objective personality tests are well suited for research investigating the factorial stability or universality of various *structure of personality* models across cultures, as well as describing salient cultural differences in personality test patterns.

Objective personality tests have been heavily used and researched in the medical field, particularly in investigating the psychogenic component of problem areas such as chronic pain, ulcers, asthma, irritable bowel syndrome, premenstrual syndrome, coronary heart disease, and recently even cancer. The MMPI and other tests have also been used to predict treatment outcome for other problems commonly encountered in behavioral medicine, such as obesity, anorexia, bulimia, smoking, and sexual dysfunction. As in other areas of research, many studies focus on describing mean group profiles in an attempt to describe *the* personality pattern characteristic of such individuals, whether it be cocaine abusers, infertile women, Hare Krishna devotees, head trauma patients, deep sea divers, shoplifters, or medical students (among many others). In the medical and psychiatric fields, such studies often suffer from the lack of empirical support for their assumption that these personality patterns are risk factors for the particular problem syndrome rather than the result of it; more prospective studies are needed to address this issue. While group profiles are useful in contrasting different diagnostic groups to search for salient differences, group profiles tend to hide the large individual variation in personality types within particular diagnostic labels, ethnic groups, occupations, or other defined groups of individuals.

Fortunately, as in the alcohol and drug abuse literature, there is a growing trend to take individual differences into account by searching for distinctive personality subgroups within an overall diagnostic label. The latter approach, used in applications such as predicting surgery outcome for back pain patients, tries to identify subgroups of patients suffering from a particular physical disorder who share a common etiology or a common problem constellation. The hope is that particular patient subgroups might benefit differentially from treatment packages different from those most useful for other patients with another pattern of contributing factors. Although studies are still being conducted to discriminate *functional* medical problems from *organic* ones, the majority consensus these days is that personality instruments are not very useful in making such a black-and-white distinction. They may be more useful in describing a constellation of personality factors, coping styles, psychiatric problems, and physical complaints with implications for treatment planning.

Research on MMPI use in forensic settings has burgeoned in the last few years. Typological systems and special scales have been developed to predict recidivism, assess potential aggressiveness, and judge the legal competence of felons. Prediction of institutional adjustment using the MMPI-based Megargee classification of personality types (Megargee & Bohn, 1977) has received some further validational work, and other scales and configural rules have been researched as descriptors of major subtypes within the "criminal personality." In a somewhat related area of relevance to public safety, there is a growing body of research on the utility of personality tests in screening for highly stressful, highly responsible occupations such as police officer or airline pilot.

Efficiency and brevity of assessment are always of concern to practitioners and researchers. The literature reflects constant development of new instruments, or new scales from old instruments, designed to improve assessment of a particular area of interest. As Evans (1983) noted in his review of trends in the research published by the *Journal of Personality Assessment* over a 15-year span, the majority of these procedures disappear from sight after an initial pub-

lished description. A particularly popular area for many years has been the effort to shorten the MMPI into a more efficient form. Several short forms continue to be used as criterion measures of psychological functioning, despite the general lack of external validity data justifying their use as equivalent forms of the MMPI (see cautions against short form use by Hoffman & Butcher, 1975; Streiner & Miller, 1986). Most of the studies we examined still evaluated the validity of short forms through correlations with long form scores, rather than looking at whether similar clinical decisions would be made based on the short versus long forms. Increasingly, clinical procedures are being streamlined by using computerized forms of many objective personality tests. The literature exploring the validity of such automated systems is certainly on the rise, but still all too scanty. The most extensive research program to evaluate computer-based interpretive reports has been conducted by Eyde, Kowal, and Fishburne (1987). Their article should serve as a model for future validity research.

Despite the age-old call for shorter, faster methods of assessment, the past few years have shown an interesting increase in studies advocating the combination of two or more personality instruments (such as the MMPI and MCMI) to provide a more thorough and detailed assessment than might be available from either test alone. If this trend truly reflects the needs of clinicians and their willingness to invest more time and energy to obtain greater accuracy of assessment, criticisms of the length of tests such as the MMPI may not be representative of most practitioners' views.

The research and clinical areas mentioned above provide a sample of the broad and varied current uses of objective personality instruments. Many other more specific studies were reported, and other topic areas, such as adolescent adjustment, family problems, relationship of personality test indices to other psychometric test data, prediction of response to psychopharmacologic treatment, custody evaluation, etc., are too numerous to explore in this brief survey.

Status of the MMPI

In the 10 years since Butcher and Owen (1978) published their survey of objective personality assessment, the MMPI has continued to be the most widely used and researched personality assessment instrument. Butcher and Owen found that 84% of the articles published between 1972 and 1977 employing a major clinical objective personality test utilized a 40-year-old inventory: the MMPI. Butcher and Keller (1984) found the same pattern was maintained between 1978 and 1982: 87% of the articles they surveyed utilized the MMPI. In both reviews, the 16PF came in a not-very-close second, and the newer DPI and MCMI had extremely little published research to validate their practical utility, beyond the original empirical studies the authors had utilized in constructing the tests.

In reviewing the literature since Butcher and Owen's survey, we found that the MMPI continues its strong domination of the published literature. This pattern varies a bit by population studied and focus of research; tests such as the 16PF and PRF are used most widely in normal populations and in more basic research on personality structure or descriptions of the intercorrelations of various personality measures. They may be viewed as more elegant instruments, both psychometrically and theoretically, than the MMPI, and their use in academic research probably reflects this. However, in clinical settings the extensive empirical validation literature on the MMPI is continuing to make it the most popular instrument. For practical decision-making applications, such validity data are essential, and are still largely missing on other objective personality instruments. Fortunately, in the past few years there has been a growing body of clinical research literature exploring the validity and utility of the MCMI, with some researchers finding it to be a very useful alternative or supplement to the MMPI in their settings. It is to be hoped that this trend will continue, with other instruments as well as the MCMI, so that psychologists will be able to make informed choices of appropriate assessment devices to be used for their particular purposes.

THE MMPI: AN EXAMPLE OF TEST EVOLUTION

The discussion above shows clearly that despite new developments in test construction that may hold promise for the future, the contemporary status of objective personality assessment is still largely synonymous with the contemporary status of the MMPI. There are several hundred new publications a year on the MMPI and the number of clinical applications involving the test appears to be growing substantially. Accordingly, we will discuss this cornerstone of contemporary objective personality measurement in more detail.

The MMPI, in one instrument, provides an example of the process of objective test development and the evolution of scale development and interpretive techniques over time. The story is largely one of adapting a preexisting instrument to the ever-changing needs of clinicians and researchers during the past four decades, a process which is continuing with the current development of the revised MMPI (MMPI-2). It is to the original authors' credit that this 45-year-old instrument has continued to prove useful for so many applications so far removed from its original purpose. Although the MMPI is classified as an empirically keyed test, the next sections make it clear that MMPI use has spanned all methods of objective test development and thus provides a good example of the practical integration of many approaches to objective personality assessment.

Original Item and Scale Development

In 1940 Hathaway and McKinley published the first of a number of articles detailing the development of a new paper-and-pencil personality test, the Minnesota Multiphasic Personality Inventory (MMPI), devised as an objective means of obtaining clinical-diagnostic information about clients in psychiatric and general medical settings (see Dahlstrom & Dahlstrom, 1980, for a collection of historically important articles on the MMPI). For the most part, scale development for the MMPI was empirical. The items selected for the scales were obtained, not according to some theoretical plan of the test developers, but through empirical item-analysis procedures designed to maximize discrimination between various predefined psychiatric groups versus "normal" individuals. The collection of scales was chosen to assess clinically relevant syndromes such as depression, hypochondriasis, sociopathy, schizophrenia, and hypomania, and to aid the clinician in arriving at a clinical diagnosis based on the diagnostic categories being used in the 1940s.

Unlike most personality inventories in existence at the time, the final selection of items for the MMPI scales followed a stringent empirical approach. However, the overall design of the MMPI cannot be called purely empirical, as selection of the original item pool was guided by preexisting ideas of appropriate content areas and themes relevant to psychiatric classification. The test authors initially collected over 1,000 items from a variety of sources such as clinical cases, textbooks, and previously published tests. These items were reduced to a pool of 504 items by eliminating those that the authors felt were redundant or were not considered promising for other reasons (again not a particularly empirical method). This reduced item pool was then administered to a group of individuals referred to as the "Minnesota normals": mostly visitors to the University of Minnesota Hospital not presently under a doctor's care, augmented by a group of students attending conferences at the university, some medical patients, and a group of Work Progress Administration workers. The clinical scales were developed by determining which items differentiated the "normal" group from various clinical reference groups. These reference groups were composed of University of Minnesota Hospital psychiatric patients who had been diagnosed as having the symptoms or the problems defined by the diagnostic categories in use at the time, for example hysteria or psychasthenia. Care was taken in the accumulation of clinical cases to assure that the clinical problems within each group were homogeneous and clearly defined by the diagnostic criteria used in the study. Eight clinical scales were constructed by this method of contrasted groups: *Hs* (Hypochondriasis), *D* (Depression), *Hy* (Hysteria), *Pd* (Psychopathic Deviate), *Pa* (Paranoia), *Pt* (Psychasthenia), *Sc* (Schizophrenia), and *Ma* (Hypomania).

Hathaway and McKinley also departed from standard practice at the time by building several validity checks into their instrument from the beginning. Self-report instruments are notoriously vulnerable to invalidating effects of the subject's test-taking style or attitude, and almost every contemporary test includes one or more validity checks. In the MMPI, the "cannot say" scale is simply the total number of items which were not answered or were answered in both directions; too high a score lowers the scores on all scales and calls the validity of the test into question. The *L* scale (for "lie") was a rationally constructed scale, composed of items designed to tap an individual's unwillingness to admit to even commonly held, minor faults. The *F* scale measures deviance of responses compared to the normative sample; it is composed of items which less than 10% of the "Minnesota normals" answered in the keyed direction. Finally, the *K* scale was designed to identify subtle clinical defensiveness. It is the only validity scale in the MMPI to have been constructed empirically by the method of contrasted groups. A group of people with known psychopathology but normal MMPI profiles was compared with a group of nonpatients with normal profiles. The resulting *K* scale was later used to develop correction factors for defensiveness for several of the clinical scales (McKinley, Hathaway, & Meehl, 1948).

Two extra clinical scales, *Mf* (Masculinity/Femininity) and *Si* (Social Introversion), were added to the MMPI later. The *Si* scale was developed by Drake (1946) by contrasting socially participative female college students with other college women who were not involved in many extracurricular activ-

ities. *Mf* was originally designed to distinguish between homosexual and heterosexual men, but few items were found to work for this purpose. Hathaway and McKinely broadened this scale by adding items which distinguished between heterosexual men and women, and also added new items from a sex differences scale designed by Terman and Miles (1936). The addition of this scale brought the total number of items in the full MMPI to 550. (The total number of items in the Group Form booklet of the MMPI is 566, due to the inclusion of 16 repeated items. Although many clinicians and patients assume that these items represent an extra validity check, they were actually included only because their placement on the form made scoring simpler.)

Administration and Scoring of the MMPI

In most situations the MMPI is administered with a reusable booklet form (either the Group Form or Form R) and an answer sheet which can be either hand scored or machine scored. Other administration formats are available for special circumstances and include the original box form (each item is printed on a card and the subject is required to sort the cards into groups according to whether the items are true or false), audio forms for those with reading difficulties, foreign translations, and computer-administered versions. In the latter category, Richards, Fine, Wilson, and Rogers (1983) have even described "a voice recognition system developed to allow patients with no hand function to take the MMPI independently" (p. 167).

Raw scores for each MMPI scale are found by counting the number of item responses made in the keyed direction for that scale. Often several of the raw scores are "corrected" for defensiveness with a proportion of the K-scale value which has become a standard part of MMPI scoring procedure. Raw scores are then transformed into T-score units when they are plotted on the profile sheet (See Figure 16.1 for a sample MMPI-2 profile. The raw scores are noted across the bottom). Traditional MMPI T-scores are non-normalized standard scores with a mean of 50 and standard deviation of 10, based on the score distributions of the "Minnesota normal" sample. T-scores can also be computed using other normative groups as the reference population. The MMPI-2 incorporates a new nationally representative normative sample, and T-scores are computed using a new "uniform" transformation procedure which makes score levels more easily comparable across scales (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). T-score conversion tables also have been published based on norms for specific age groups, ethnic or racial groups, or for countries other than the United States (cf. Butcher & Pancheri, 1976; Colligan, Osborne, Swenson, & Offord, 1984; Dahlstrom, Lachar, & Dahlstrom, 1986; Dahlstrom, Welsh, & Dahlstrom, 1972). Scale elevations above two standard deviations from the mean (a score elevated above a T-score of 70) are generally considered clinically significant on the original MMPI. On the revised MMPI-2, the new normative group and T-scores lower this "clinically significant" cutoff to 65 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). (It should be noted that the normative tables provided by Colligan et al. use normalized T-scores rather than the traditional non-normalized standard scores, and thus are not easily interpretable using the extensive body of research based on the latter (Hsu, 1984; Miller & Streiner, 1986). Colligan and colleagues have been publishing research which will aid in direct interpretation of their normalized T-scores; in the meantime, their provision of a method to convert normalized T-scores back to linear scores may help in applying standard interpretive rules to their data [Osborne & Colligan, 1986].)

As with many other tests, computer scoring of MMPI records is becoming increasingly popular. There are several approaches to computer scoring of the subject's responses. Printed answer sheets can be key punched onto data cards or optically scanned and stored on a magnetic data tape, then entered into a scoring program. Alternatively, subject responses can be scored as the subject is taking the test interactively. Computer scoring, interpretation, and administration of objective personality tests such as the MMPI will be discussed in more detail later in this chapter and in Hanaker and Fowler's chapter in this handbook.

Subsequent Developments

Initially it was thought that simple scale elevations would directly reflect psychopathology measured by the individual MMPI scales. For example, depressed patients would be expected to show elevations on the Depression scale but not on the Schizophrenia scale. However, it was soon apparent that the MMPI did not perform very well as a means of classifying patients into strict diagnostic groups. Patients with elevations on one scale tended also to have elevations on other scales, some psychiatric patients did not have significant elevations on the appropriate scales, and some "normal" individuals did have elevations on one or more scales. As Graham (1987) notes, over-time

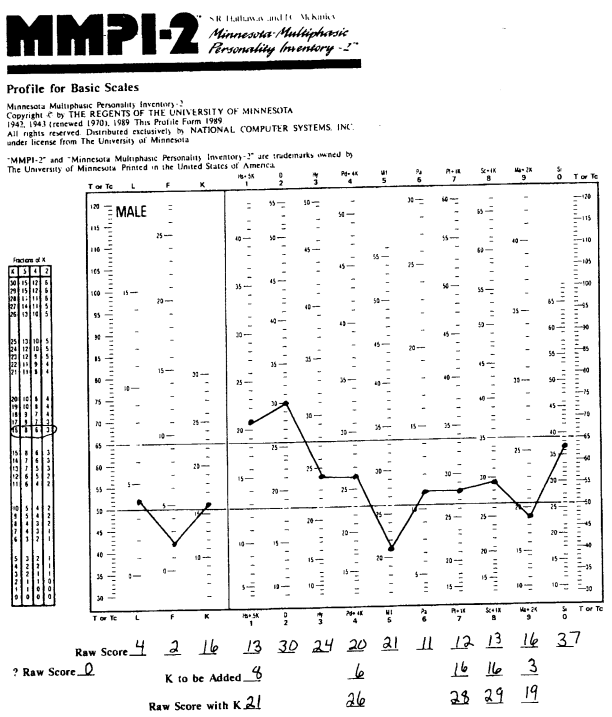


Figure 16.1. Sample MMPI-2 profile form. Profile Minnesota Multiphasic Personality Inventory-2. Copyright © by the Regents of the University of Minnesota 1942, 1943 (renewed 1970), 1989. This profile from 1989. Reproduced by permission of the publisher, the University of Minnesota Press.

interpretation of the MMPI has moved away from treating each scale as a measure of a known psychopathological category, instead of adopting a purely empirical approach:

> the MMPI is currently used in a way quite different from the way in which it was originally intended to be used. It is assumed that the clinical scales are measuring something, because reliable differences in scores are found among individuals from different reference groups. The new approach treats each MMPI scale as an unknown and, through experience and empirical research, the correlates of each scale are identified. When a person obtains a score on a particular scale, the clinician attributes to that person the characteristics and behaviors that through previous research and experience have been identified for other individuals with similar scores on that scale. . . . It is this new behavioral description approach to the utilization of the MMPI in everyday practice that has led to the instrument's great popularity among practicing clinicians. (pp. 7–8)

As a result of this changing emphasis in MMPI interpretation, the 10 clinical scales are now usually referred to by numbers rather than by their original names (see Table 16.1).

Configural Interpretation.

The empirical correlate approach to interpretation of individual scales has also been expanded to interpretation of the overall pattern of scale scores obtained by an individual. Partly because of the intercorrelation among scales, but also presumably because of the overlap among characteristics of clinical syndromes, it was often found that several MMPI scales would be elevated together. Interpretation of a particular scale might vary depending on the relative elevations of

Table 16.1. Summary of Correlates of the MMPI Validity, Clinical, and Selected Special Scales

MMPI SCALE	SCALE CORRELATES
Validity Scales	
? (Cannot Say)	The number of items not answered or answered in both directions. A defensive or invalid profile with possible attenuation of scale scores is suggested if the ? raw score is 30 or more.
L (Lie)	Measures a rather unsophisticated or self-consciously "virtuous" test-taking attitude. Elevated scores suggest that the individual is presenting himself or herself in an overly positive light, attempting to create an unrealistically favorable view of his or her adjustment.
F (Infrequency)	Items on this scale are answered very infrequently by most people. A high score suggests not only an exaggerated pattern of symptom checking that is inconsistent with accurate self-appraisal, but also confusion, disorganization, or actual faking of mental illness.
K (Defensiveness)	High scores reflect an uncooperative attitude and an unwillingness or reluctance to disclose personal information or problems. Low scores suggest openness and frankness. *K* is positively correlated with intelligence and educational level, which should be taken into account in interpretation.
Clinical Scales	
1 (*Hs*, Hypochondriasis)	High scorers present numerous vague physical problems that tend to be chronic. They are generally unhappy, self-centered, whiney, complaining, hostile, and attention-seeking in their behavior.
2 (*D*, Depression)	Reflects depressed mood, low self-esteem, and feelings of inadequacy. High scorers are described as moody, despondent, pessimistic, distressed, high-strung, lethargic, over-controlled, and guilt-prone. Elevations may reflect great discomfort and need for change or symptomatic relief.
3 (*Hy*, Hysteria)	High scorers tend to rely on neurotic defenses such as denial and repression to deal with stress. They tend to be dependent, naive, outgoing, infantile, and narcissistic. Their interpersonal relations are often disrupted, and they show little insight into problems. High scorers show little interest in psychological processes and interpret psychological problems as physical ones. High levels of stress are often accompanied by the development of physical symptoms.
4 (*Pd*, Psychopathic Deviate)	Measures antisocial behavior, such as rebelliousness, disrupted family relations, impulsiveness, difficulties with school or work, legal involvement, and alcohol or drug abuse. Personality disorders are likely among high scorers: they are outgoing, sociable, and likeable as well as deceptive, manipulative, hedonistic, exhibitionistic, inclined toward poor judgment, unreliable, immature, hostile and aggressive. High scores usually reflect long-standing character problems that are highly resistant to change.
5 (*Mf*, Masculinity-Femininity)	High-scoring men are described as sensitive, aesthetic, passive, or feminine. They may show conflicts over sexual identity and low heterosexual drive. Low scoring men are viewed as masculine, aggressive, crude, adventurous, reckless, practical, and having narrow interests. Because the direction of scoring is reversed, high-scoring women are seen as masculine, rough, aggressive, self-confident, unemotional, and insensitive. Low-scoring women are viewed as passive, yielding, complaining, fault-finding, idealistic, and sensitive.
6 (*Pa*, Paranoia)	Elevations on this scale are often associated with being suspicious, aloof, shrewd, guarded, worrying, and overly sensitive. High scorers may project or externalize blame and harbor grudges against others. They are generally hostile and argumentative.
7 (*Pt*, Psychasthenia)	High scorers are tense, anxious, ruminative, obsessional, phobic, and rigid. They frequently are self-condemning and guilt prone, feel inferior and inadequate, overintellectualize and rationalize problems, and resist psychological interpretations.

Table 16.1. Summary of Correlates of the MMPI Validity, Clinical, and Selected Special Scales (*Continued*)

MMPI SCALE	SCALE CORRELATES
8 (*Sc*, Schizophrenia)	High scorers have an unconventional, schizoid life-style. They are withdrawn, shy, moody, and feel inadequate, tense, and confused. They may have unusual or strange thoughts, poor judgment, and erratic moods. Very high scorers may evince poor reality contact, bizarre sensory experiences, delusions, and hallucinations. They are generally uninformed and have poor problem-solving skills.
9 (*Ma*, Hypomania)	High scorers are viewed as sociable, outgoing, impulsive, overly energetic, and optimistic. They have liberal moral views, are flighty, may drink excessively, are grandiose, irritable, impatient, and rarely "follow through" with their plans. They are manipulative and exaggerate their self-worth. Very high scorers may show affective disorder, bizarre behavior, erratic mood, impulsive behavior, and delusions.
0 (*Si*, Social Introversion)	High scorers are viewed as introverted, shy, withdrawn, socially reserved, submissive, overcontrolled, lethargic, conventional, tense, inflexible, and guilt-prone. Low scorers are extroverted, outgoing, gregarious, expressive, aggressive, talkative, impulsive, uninhibited, spontaneous, manipulative, opportunistic, and insincere in social relations.
Special Scales	
A (Anxiety)	This scale defines the first and the largest factor dimension in the MMPI. It measures general maladjustment or emotional upset. High scores reflect anxiety, tension, lack of efficiency, and open expression of numerous psychological complaints.
R (Repression)	This factor scale relates to reliance on denial and repression. High scores reflect uninsightful, overcontrolled, and inhibited behavior. These individuals tend to avoid problems and appear overly conventional.
Es (Ego Strength)	This scale was originally developed to predict successful response to psychotherapy. Subsequent research has shown it to be an indicator of a person's overall level of functioning. High scores reflect effective functioning and the ability to withstand stress. Such individuals tend to have psychological resources that will help them to cope with problems.
MAC (MacAndrew Addiction)	This scale was developed to distinguish alcoholic from non-alcoholic psychiatric patients. High scores are also associated with other addictive problems such as drug abuse and pathological gambling; *MAC* serves as a measure of addiction-proneness.

Adapted from University of Minnesota Press (1984), *User's Guide for the Minnesota Report: Personnel Selection System.* Minneapolis, MN: Author.

other scales in the profile. Clinicians found that individuals with common MMPI profile patterns were often similar in terms of manifest problems, symptoms, and personality characteristics. The pattern or *configural approach* to MMPI interpretation came to be the preferred interpretive strategy (Halbower, 1955; Meehl, 1954). Several subsequent research programs have delineated actuarially based empirical descriptions of various MMPI patterns referred to as *code types*. For example, the individual whose profile is shown in Figure 16.1 has highest scale elevations on scales 2 (originally Depression) and 1 (originally Hypocondriasis), and therefore would be classified as a 2–1. Correlates of this pattern could be looked up in one of many so-called cookbooks listing research results (Dahlstrom, Welsh, & Dahlstrom, 1972; Gilberstadt & Duker, 1965; Graham, 1987; Greene, 1980; Lachar, 1974; Lewandowski & Graham, 1972; Marks, Seeman, & Haller, 1974; Nelson & Marks, 1985). This work serves as the empirical basis of objective MMPI profile interpretation.

Another strategy for MMPI interpretation which combines information from several scales is the use of *objective rules* or *objective indices* to arrive at clinical decisions. This interpretive approach can use either actuarially derived discrimination rules such as the

Goldberg Index (Goldberg, 1965) or clinically compiled classification rules such as the Henrichs Rules (Henrichs, 1964, 1966) or the Meehl-Dahlstrom Rules (Meehl & Dahlstrom, 1960), designed to distinguish between neurotics and psychotics. In this approach the test interpreter, after scoring the record and drawing the profile, computes the index or classifies the profile according to the decision rules. For example, the Goldberg Index is computed by adding scales *L, Pa,* and *Sc* together and then subtracting scales *Hy* and *Pt* to get an index of the likelihood the patient is psychotic. As another example, if classification rules for combining scale scores confirm that a correctional inmate meets Megargee's "I" classification in his system for grouping criminal offenders (Megargee & Bohn, 1977), then the correlates and predictions relevant to that felony group would be applied to the case. An advantage of the objective classification approach is that it can readily be applied by a clerical worker or by a computer. Often, in fact, the rules become so complicated that automation is a distinct advantage in utilizing the research on such configural combinations.

Content-based Interpretation.

Although early MMPI research developments and clinical applications emphasized empirical correlates and objective interpretation of profiles, there were also advocates of *content-based interpretation* (Grayson, 1951; Harris & Lingoes, 1968; Koss & Butcher, 1973; Wiggins, 1969). Content interpretation takes us back to the rational method of test construction and interpretation; it assumes that the subject, in answering test items, is reacting and responding openly and directly to the item meaning and content. Consequently, the content of MMPI items might represent an important source of information that is not available through empirical test interpretation procedures. Several approaches to interpretation of MMPI content have been developed.

1. The *factor-analytic* approach. Several researchers have attempted to develop keys for factor dimensions within either the entire MMPI or within specific scales. Welsh (1956) and Block (1965) developed factor scales by intercorrelating the standard MMPI scales; their two primary factors are now scored by many clinicians and researchers as "*A*" (anxiety) and "*R*" (repression). Content-consistent subscales for *Mf* and *Si* were developed by Serkownek (see Schuerger, Foerstner, Serkownek, & Ritz, 1987) based on factor analyses of the items in these two scales. Others (cf. Comrey, 1957; Costa, Zonderman, McCrae, & Williams, 1985; Johnson, Butcher, Null, & Johnson, 1984; Stein, 1968) attempted to factor analyze the MMPI using item intercorrelations to arrive at homogeneous factor scales. While the factor analytic approach is not specifically oriented toward obtaining "pure" content dimensions, the research has produced a number of scales with homogeneous content.

2. The *rational* approach to content analysis. Several efforts to develop content scales by rational analysis have been attempted. One set of content scales, the Harris-Lingoes subscales for the MMPI (Harris & Lingoes, 1968), has become widely used in clinical and computerized MMPI interpretation. The Harris-Lingoes subscales are subsets of MMPI items on several MMPI scales: *D, Hy, Pd, Pa, Sc,* and *Ma*. The subscales were developed by reading through the items and placing "similar" items into content groups. A more systematic approach to developing content scales (Wiggins, 1969) included all of the MMPI items in the pool of potential contents and developed 13 homogeneous content scales for the MMPI. These scales are *HEA* (Health), *DEP* (Depression), *ORG* (Organic Symptoms), *FAM* (Family Problems), *AUT* (Authority Conflict), *FEM* (Feminine Interests), *REL* (Religious Fundamentalism), *HOS* (Manifest Hostility), *MOR* (Poor Morale), *PHO* (Phobias), *PSY* (Psychoticism), *HYP* (Hypomania), and *SOC* (Social Maladjustment). High scores on the content scales reflect high degrees of endorsement of the themes represented in the content scales. However, there have been relatively few studies establishing empirical correlates for the content scales. Thus, their interpretation must be based primarily upon rational or intuitive analysis of the scales by examining the "meanings" of the content dimensions reflected in the elevated scale.

3. The *critical item* approach to content interpretation. This approach assumes that individuals disclose personal problems through their response to individual items and suggests that certain item groups are more significant than others for reflecting important problem areas. Grayson (1951) first recommended the use of specific MMPI items as important signs of psychopathology. The items he suggested as "critical" for the assessment of psychopathology have been referred to as the "MMPI Critical Items," and have received widespread use despite the lack of empirical verification of their validity as problem indices. More recently Koss and Butcher (1973; Koss, Butcher & Hoffman, 1976), and Lachar and Wrobel (1979) have published empirically based critical items. The use of such critical items for fleshing out MMPI interpretations has gained wide acceptance. Although there are some problems with this interpretive approach (Koss, 1980), particularly the low reliability of single items, this content-based procedure may add considerably to the empirical method of test interpretation.

Interpretation of special scales.

Current research and practice with the MMPI rely primarily on use of the validity scales, the clinical scales, the content scales and a number of special or experimental scales. Several hundred additional scales have been developed for the MMPI, primarily for specific purposes, such as measuring personality or behavioral characteristics like dominance or prejudice, or for predicting various symptom patterns such as drug abuse or chronic pain syndrome. Most of these additional scales have not attained wide use or acceptability. However, a few of the special scales have been widely researched and used clinically. Some of these experimental scales are the Ego Strength Scale (*Es*), a measure of tolerance for stress or ego strength (Barron, 1953); the Welsh Anxiety Scale (*A*), a factor scale designed as a measure of general maladjustment or anxiety (Welsh, 1956); the Welsh Repression Scale (*R*), another factor scale designed as a measure of overcontrol; and the MacAndrew Alcoholism Scale (*Mac*), a measure of an individual's proneness to addiction (MacAndrew, 1965). Table 16.1 provides a brief description of the basic validity and clinical scales for the MMPI, along with these four special scales. More detailed description of the many MMPI scales and scale combinations can be found in a number of useful texts (cf. Dahlstrom, Welsh, & Dahlstrom, 1972, 1975; Duckworth & Anderson, 1986; Graham, 1987; Greene, 1980).

It should be apparent that interpretation of the MMPI is a complicated process that requires professional training and access to a large amount of actuarial data. Later in this chapter we will illustrate the integration and application of all these interpretive methods to the case in Figure 16.1, through a report generated by the second author's computerized interpretive system.

Revision and Restandardization of the MMPI

In their overview of recent developments in the use of the MMPI, Butcher and Owen (1978) noted a number of problems and criticisms of the MMPI. Some of these problems involved the MMPI itself and centered around the apparent need for a revision and restandardization of the inventory; other problems involved the relative inactivity on the part of the test distributor in keeping up with the existing MMPI technology and in failing to provide necessary interpretive materials that had been developed for the MMPI. The problems cited included the need to update and broaden the MMPI item pool; the need for a new standardization of normal responses on a broader, more representative contemporary normative sample; and the need for more flexibility and willingness on the part of the test distributor to provide relevant test materials.

Since that review by Butcher and Owen a number of changes have taken place in the publication and distribution of the MMPI that may serve to rectify some of these problems. The major administrative change was that the University of Minnesota Press, the MMPI copyright holder, resumed responsibility for research and development of the instrument. The University Press contracted with National Computer Systems of Minneapolis to be the licensed distributor of the MMPI and to develop new profile sheets, scoring sheets, and other MMPI related materials to enhance clinical use of the test. The profile form illustrated in Figure 16.1 is an example of these updated materials.

In 1982 the University of Minnesota Press initiated a major research effort to revise, update, and restandardize the MMPI (Butcher, Graham, Dahlstrom, Tellegen, & Kaemmer, 1989). Some of the goals set for the redevelopment of the MMPI were the following:

1. To maintain the integrity of the existing validity, clinical and widely used special scales of the test.

2. To revise and reword the language of some of the existing items that were out of date, awkward, sexist, etc. Over the years language use has changed, and the content of some of the MMPI items has become antiquated.

3. To broaden the item pool to include other contents not now represented—for example, treatment compliance, amenability to change, relationship problems, work attitudes, etc.

4. To develop new, up-to-date norms for the MMPI. The original "Minnesota normals" represented a small, regional, somewhat parochial sample of adults living 40 years ago. Regional normative samples collected in North Carolina by Diehl (1977) and in southern Minnesota by Colligan et al. (1984) have indicated that the old norms are probably not representative of response patterns of contemporary "normal" people, but no nationwide standardization had ever been undertaken.

5. To include separate forms of the MMPI for adults and adolescents. New items were included for the adolescent form of the MMPI that are specific to problems of adolescents.

The MMPI project committee (James Butcher of the University of Minnesota, Grant Dahlstrom of the University of North Carolina, Jack Graham of Kent State University, and Auke Tellegen of the University of Minnesota) chose to conduct a program of research that was both conservative and expansive in scope: goals were to maintain the integrity of the original instrument while expanding its range of coverage, utility, and acceptability to clients.

To insure continuity with the original MMPI and its extensive research base, the committee decided to include the entire existing MMPI item pool (550 items) in the experimental booklet so that the original items and scales could be studied in modern samples of normals. It was possible for users involved in the restandardization research to continue to score and interpret the original MMPI scales while collecting responses on the new instrument. The 16 repeated items in the original MMPI were deleted and replaced with new items described below. This change would not affect scoring of the basic MMPI scales, as the repetitions were not scored in the original MMPI. About 14% of the original items were changed because of dated language or content, sexist or otherwise objectionable wording, or awkward grammar. In a few instances where items were so out of date as to be meaningless, new items were substituted. However, for comparison purposes, the original item was retained (along with the rewritten version) in the experimental booklets. Analyses presented by Ben-Porath and Butcher (1988) revealed that out of 82 rewritten items, only nine showed significant differences in endorsement percentage when compared to their original version, and none of these differences held across both sexes. Rewriting the items did not change any item-scale correlations significantly, and so had no real effect on the psychometric characteristics of the MMPI.

In addition to retaining the original items, some in updated form, the experimental MMPI booklet was expanded by the addition of 154 new items designed to address problem areas not well represented in the original version of the test. These additional items were selected rationally, through a broad sampling of views of MMPI experts as to which content domains needed further coverage. The separate adolescent form of the experimental MMPI booklet also contained the original 550 MMPI items, 50 new items dealing with treatment amenability, and 104 new items designed to provide better coverage of concerns and problems specific to adolescents.

The official MMPI-2 restandardization project involved an extensive collection of new MMPI, biographical, demographic, and life event data on a *national* normative sample of adults (N = 2,600) and adolescents (still being tested). The new subject population for the MMPI-2 was obtained through a sampling of normal volunteers from several regions of the United States: Minnesota, Ohio, North Carolina, Pennsylvania, Washington, Virginia, and California. Efforts were made to obtain a sample representative of the U.S. population by matching sample characteristics to major demographic characteristics reported in the 1980 census. For a large subsample of the adult population, both members of married couples were tested and were asked to fill out behavioral and personality ratings of each other as well as assessing their marital relationship. The extra forms used in the national standardization provided descriptive data and MMPI response correlates that were not available for the original Minnesota normative sample.

The new norms for the MMPI obtained from this national sample will be eased into use in a conservative fashion. Clinicians and researchers will be provided with materials to allow direct comparison between old and new norms to insure continuity of interpretive strategies and research data. The MMPI-2 manual presents tables facilitating such comparisons (Butcher et al., 1989). Although the MMPI-2 booklet contains some of the new, experimental items which clinical studies have shown to be promising, the items contained in the original MMPI validity and clinical scales will be placed at the beginning of the booklet for those individuals who wish to use only the original scales.

Over time, however, it is likely that users will find many of the new inclusions in the MMPI-2 to be as useful or even more useful than the original MMPI scales. For example, content interpretation of the MMPI has been expanded and improved by the development of several new MMPI-2 content scales which include the dimensions represented by new items in the MMPI-2 item pool (See Table 16-2: Butcher, Graham, Williams, & Ben-Porath, 1989). These scales were developed by a multistage, multimethod approach starting with rational item groupings and then proceeding through statistical item-selection techniques to improve individual scale homogeneity and reduce scale intercorrelations. A number of studies aimed at testing and validating these scales against contemporary clinical populations have been conducted (Graham & Butcher, 1988; Keller & Butcher, in press). Some of the new items have already been shown to improve the discrimination of existing scales as well as to provide new content for assessing other problem areas. We will likely find, as new clinical

Table 16.2. MMPI-2 Content Scales

SCALE	DESCRIPTION OF CONTENT AND CORRELATES
ANX (*Anxiety*)	General symptoms of anxiety and tension; sleep and concentration problems; somatic correlates of anxiety; excessive worrying; difficulty making decisions; willingness to admit to these problems.
FRS (*Fears*)	Many specific fears and phobias including animals, high places, insects, blood, fire, storms, water, the dark, being indoors, dirt, etc.
OBS (*Obsessiveness*)	Excessive rumination, difficulty making decisions, compulsive behaviors, rigidity, feelings of being overwhelmed.
DEP (*Depression*)	Depressive thoughts, anhedonia, feelings of hopelessness and uncertainty, possible suicidal thoughts.
HEA (*Health Concerns*)	Many physical symptoms across several body systems: gastrointestinal, neurological, sensory, cardiovascular, dermatological, and respiratory. Reports of pain and of general worries about health.
BIZ (*Bizarre Mentation*)	Psychotic thought processes, auditory, visual, or olfactory hallucinations, paranoid ideation, delusions.
ANG (*Anger*)	Anger control problems, irritability, impatience, loss of control, past or potential abusiveness.
CYN (*Cynicism*)	Misanthropic beliefs, negative expectations about the motives of others, generalized distrust.
ASP (*Antisocial Practices*)	Cynical attitudes; problem behaviors; trouble with the law, stealing, belief in getting around rules and laws for personal gain.
TPA (*Type A*)	Hard-driving, work-oriented behavior; impatience; irritability; annoyance; feelings of time pressure; interpersonally overbearing.
LSE (*Low Self Esteem*)	Low self-worth; overwhelming feelings of being unlikable, unimportant, unattractive, useless, etc.
SOD (*Social Discomfort*)	Uneasiness around others; shyness; preference for being alone.
FAM (*Family Problems*)	Family discord, possible abuse in childhood, lack of love and affection in the family or marriage, feelings of hate for family members.
WRK (*Work Interference*)	Behaviors or attitudes likely to interfere with work performance, such as low self esteem, obsessiveness, tension, poor decision making, lack of family support, negative attitudes towards career or coworkers.
TRT (*Negative Treatment Indicators*)	Negative attitudes toward doctors and mental health treatment. Preference for giving up rather than attempting change. Discomfort discussing any personal concerns.

(Adapted from Butcher, Graham, Williams, and Ben-Porath, 1989)

populations are studied, that new items and scale-development strategies will add to and sometimes even replace traditional approaches to MMPI use.

Comments

MMPI scale development and interpretive strategies have increasingly incorporated methods, such as content analysis and factor analysis, that range far from the test's original empirical-criterion-keyed roots. However, the rationale for inclusion of these new methods and for revision of such tried-and-true personality tests remains the same: to provide a more efficient, accurate, reliable, and valid tool to aid psychologists in research and practical decision making. Many of the new developments with the MMPI, as well as with other objective personality tests, would not have been possible without the aid of another tool: the computer. The next sections will focus on the increasingly integral role of automation in the contemporary and future status of objective personality assessment technology.

AUTOMATED OBJECTIVE PERSONALITY ASSESSMENT

In 1981, Greist and Klein reported that although 70% of state departments of mental health and 50% of community mental health centers used computers, very few of these employed them for clinical uses. Instead, computer technology had been introduced to meet management needs for information on resources, materials, facilities, staff, finances, patient population characteristics, and services provided. Such administrative and fiscal operations traditionally have been viewed as more amenable to automation than the less understood, less structured, and less specific techniques of clinical practice. Nevertheless, the literature of the past few years clearly demonstrates the growing use of computers as tools to aid in clinical as well as administrative decision making.

Johnson, Giannetti, and Williams (1976) described the history of computer use in mental-health care delivery as a progression from (a) automated patient data systems to (b) development of automated clinical techniques to (c) the development of interventionally relevant, decision-oriented strategies. Early computer applications (such as patient data management, test scoring, and interpretive programs designed to mimic traditional clinical practice) capitalized primarily on the speed, objectivity, and enormous memory capacity of the computer. More recent innovative applications have begun to utilize the flexibility and unique adaptive capabilities of computers, modifying old assessment strategies and developing new ones.

Automated Test Scoring and Interpretation

Automated scoring of objective test answer sheets was one of the first clinical applications of computer technology, followed by computerized interpretation of standard psychological tests. Objective personality inventories are particularly amenable to automation because of their limited, structured response options and because of the standardized interpretive rules for these tests. The computer can store and access a much larger fund of interpretive literature and base-rate data than any individual clinician can master, contributing to the accuracy, objectivity, reliability, and validity of computerized reports. While any single clinician may do well to learn the norms and base rates of the client population he or she sees most often, a computer program can refer to a variety of population norms and, if programmed to do so, will always "remember" to tailor interpretive statements according to modifying demographic data such as education, marital status, and ethnicity.

Rome, Swenson, and colleagues (1962) at the Mayo Clinic developed the first computerized scoring and interpretive system for the MMPI. Since that time, their system has been expanded and many other, more complicated computerized interpretive systems have been developed, not only for the MMPI but also for instruments such as the 16PF (Karson & O'Dell, 1975, 1987), the MCMI (Millon, 1982, 1987), and even "a system of automated interpretation of coded visual-motor responses elicited by horizontally symmetrical, ambiguous, and indeterminate visual stimuli, i.e., by the ten Rorschach inkblots" (Piotrowski, 1980, p. 85; see also Exner, 1987). Hanaker and Fowler's chapter in this handbook provides a more thorough discussion of computerized interpretive systems. Reviews or discussions of some specific automated interpretive systems can be found in Adair (1978), Butcher (1978), Eyde (1985), Fowler (1985), Green (1982), Karson and O'Dell (1975, 1987), Nichols (1985), and Sundberg (1985a, 1985b). Butcher (1987) provides a comprehensive list of psychological software and services which were commercially available at the time of his review.

It should be noted that although the computer provides enormous advantages in the application of empirical scoring and interpretive rules to personality test data, currently available automated interpretations are not truly actuarial in nature. Actuarial prediction is "arrived at by some straight-forward application of an equation or table to the data . . . the defining property is that no judging or inferring or weighing is done by a skilled clinician" (Meehl, 1954, pp. 15–16). As discussed earlier, even the interpretation of the well-researched MMPI is subject to the experience and judgment of the individual clinician using it at the time; he or she must fill in the interpretive gaps left by the available "cookbooks" and modify standard interpretations as appropriate for moderating demographic variables, local base rates, or the particular purpose of the evaluation. Thus, currently available computerized interpretive systems are actually "a mixture of rules based on actuarially validated relationships and other rules culled from established clinical lore or from the expert author's personal experience with the test" (Vale & Keller, 1987, p. 72).

Unfortunately, the naive consumer may assume that a computerized interpretation is accurate simply because it is generated by a machine rather than a human. It is important to keep in mind that "the accuracy (validity) of these kinds of interpretations is dependent on the knowledge and skill of the clinician who generated the interpretive statements. The validity of these interpretations should not be assumed and needs to be demonstrated every bit as much as the validity of a test needs to be demonstrated" (Graham, 1987, p. 231). Ben-Porath and Butcher (1986) have suggested several questions that the potential user of a computerized test interpretation should consider:

> (a) To what extent has the validity of this report been studied?
> (b) To what extent does this report rely on empirical findings in generating its interpretation?
> (c) To what extent does this report incorporate all of the currently validated descriptive information?
> (d) Does the report take demographic variables into account?
> (e) Are different versions available for various referral questions?
> (f) Is the report internally consistent?
> (g) Does the report include practical suggestions?
> (h) Is the report periodically revised to reflect newly acquired information? (pp. 173–174)

The updated guidelines for computer-based tests and interpretations published by the American Psychological Association make it clear that it is the responsibility of the professional using a computerized test to evaluate its validity for the particular application in question (American Psychological Association, 1986). Validation of computerized interpretations is a complex endeavor which is generating considerable controversy and, fortunately, some new research studies at this time. For further discussion/debate of the issues, strategies, and difficulties in validation of automated interpretations, see Eyde, Kowal, and Fishburne (1987), Fowler and Butcher (1986), Matarazzo (1986a, 1986b), and Moreland (1985). It should also be noted that although there are numerous computer scoring programs available through commercial outlets for many contemporary objective personality tests, there is no assurance that these programs will generate scores or profiles that match those in the test handbooks. The user should never assume programs are error free; even the basic scores should be verified before use.

Figures 16.2 and 16.3 present a case description and sample computer-generated report for the man whose MMPI-2 profile appeared in Figure 16.1. This particular automated system, the Minnesota Report Adult Clinical System (Butcher, 1989a), illustrates a mixed actuarial and clinician-modeled interpretive program. The report is modeled on the interpretive style of James Butcher, incorporating all the strategies described earlier in this chapter: code type and individual scale interpretation, content analysis using the new MMPI-2 scales as well as traditional subscale analysis, and utilization of well-researched special scales. The program is arranged hierarchically, first checking for a match to a codetype with well-researched actuarial correlates, then proceeding to individual scale interpretation if no such match is found. The computer program tailors interpretive statements according to the subject's population (mental health outpatient or inpatient, medical, chronic pain adult correctional, or college counseling), and according to demographic data such as education, sex, and marital status, which research has shown to be modifiers of interpretive rules. (The interpretive report illustrated in Figure 16.3 was generated for a chronic pain patient setting.) Statements are modified or added based on content scale scores as well. Profiles are plotted for the standard validity and clinical scales and for the MMPI-2 content scales. Indices and decision rules appropriate to the client's setting are reported, along with a list of the Koss-Butcher and Lachar-Wrobel critical items the client endorsed. Comparison between MMPI-2 and MMPI norms is facilitated by presentation of Welsh codes for both scoring systems on the clinical profile form page.

A general computerized report such as that shown above may not always be appropriate for the particular application needed by the test user. A human clinician can, of course, modify his or her narrative style (or the extent of information provided) based on the intended audience and the questions to be addressed by the report. With increasing frequency, authors of computerized interpretations are also designing computer programs to address the decision-relevant information needed for a particular application or setting, as well as tailoring reports to fit the norms of particular client populations. The Minnesota Report Personnel Selection System, for example, is a modification of Butcher's interpretive system specifically designed to aid in screening applicants for "highly responsible positions involving public trust and safety and for occupations involving high degrees of stress" (Butcher, 1989b, p. 2).

Users can select either a narrative interpretation geared to address possible work problems, or a shorter screening report rating the applicant on five dimensions: openness to evaluation, social facility, addiction potential, stress tolerance, and overall adjustment. The latter, shown in Figure 16.4 for the same individual described in Figures 16.2 and 16.3, illustrates a modification of a test interpretation to address only those areas of most direct concern for the testing application: in this case, the suitability of this individ-

ual for the "highly responsible position" of airline pilot. Another difference between this report and the Adult Clinical System interpretation is the inclusion of "content themes" based on configural rules combining scales with similar content. The report shown in Figure 16.4 would usually include scores and profiles for the validity, clinical, content and many special scales, just as in the Adult Clinical System. These were excluded here because of their redundancy with Figure 16.3.

These reports, based on a single expert's interpretive framework, represent only a few of the possible methods for constructing a computerized test interpretation. For example, Vale, Keller, and Bentz described developing a computer narrative report by starting with a list of important questions to be addressed by the interpretation, then writing a statement library based on the combined expertise of a committee of interpretive experts (Vale, Keller, & Bentz, 1986; Vale & Keller, 1987). They, along with Fowler (1987), give some further guidelines on strategies and issues in writing a computerized test interpretation system.

Online Assessment Using Standard Clinical Techniques

Although most automated testing systems have merely scored and interpreted test forms that were hand administered, the computer is increasingly being utilized as test administrator as well. Lushene, O'Neil, and Dunn (1974) first reported an online MMPI administration, scoring, and interpretive system, and several large clinical settings now routinely use computer-administered assessment devices. Teleprocessing of test responses is possible through commercial scoring services: users can enter subjects' test responses into a remote terminal and have the data transmitted to the central scoring service via regular phone lines, receiving almost instant turnaround on the test results. Alternatively, the test can be administered directly to the subject via the terminal and her or his responses can be sent to the scoring service without the need for an intermediary data-entry step.

Erdman and colleagues reviewed 21 studies of computerized patient interviews and concluded that they are "reliable, accurate, and highly acceptable to

Case Description

Mr. S. is a 48-year-old married Caucasian man who had been employed for 30 years as a mail carrier before suffering the gradual onset of low back pain over a three- to four-year period. He worked intermittently during these years but had been away from work on disability for two years by the time of this testing, which was part of his assessment for admission to an inpatient chronic pain rehabilitation program. Mr. S.'s pain and disability were considered to be in excess of organic findings. He was admitted to a three-week rehab program consisting of physical therapy, education, vocational counseling, family counseling, therapeutic recreation, stress management training, etc. During the program, staff collected history indicating the client had previous hospitalizations for "anxiety attacks" every year or so. He would check himself into hospitals for "rest" and minor tranquilizers. The internist who evaluated him during his chronic pain program speculated about "an element of malingering" in his clinical picture. The client was also evaluated by a psychiatrist, who suggested major depression camouflaged by somatization. He noted that some of the client's disability could be traced to secondary gain and attention-seeking. He was dependent on Ativan, but was weaned off of this and started on an anti-depressant medication instead. The client was cooperative during inpatient treatment but when he went back to work afterwards, his behavior was so withdrawn and unusual that he was rehospitalized on a psychiatric unit. Staff there evaluated him and concluded that his fear of the social and work responsibilities of his job were leading to use of "bizarre psychiatric symptoms as a way of controlling the situation." The MMPI clinical and content scales helped rule out actual psychotic processes in favor of more avoidant, anxious behaviors motivated by secondary gain and social discomfort. The client basically was hoping for retirement or continued disability rather than a return to work. When this was addressed directly and the client was given a great deal of support, stress management and assertiveness training, he was able to return to work and maintain employment during the several months of followup information available, despite continued vague somatic complaints.

Figure 16.2.


```
                                    TM
                                MMPI-2
                                                        TM
MINNESOTA MULTIPHASIC PERSONALITY INVENTORY-2

By Starke R. Hathaway and J. Charnley McKinley


                                                TM
                THE MINNESOTA REPORT:
                ADULT CLINICAL SYSTEM
                 INTERPRETIVE REPORT

                By James N. Butcher


                                Client ID:  0025297
                            Report Date:  23-AUG-89
                                    Age:  48
                                    Sex:  Male
                                Setting:  Chronic Pain
                              Education:  12
                          Marital Status:  Married


        MINNESOTA MULTIPHASIC PERSONALITY INVENTORY-2
    Copyright (c) by THE REGENTS OF THE UNIVERSITY OF MINNESOTA
      1942, 1943 (renewed 1970), 1989.  This Report 1989.
                    All Rights Reserved.
    Distributed Exclusively by NATIONAL COMPUTER SYSTEMS, INC.
        Under License from The University of Minnesota.

    "MMPI-2," "Minnesota Multiphasic Personality Inventory-2," and
"The Minnesota Report" are trademarks owned by The University of Minnesota.
                Printed in the United States of America.
```

Figure 16.3. Computer-generated narrative interpretation and score profile for the individual described in Figure 16.2. Reproduced by permission of the publisher, The University of Minnesota Press.

TM page 1
MMPI-2 TM
THE MINNESOTA REPORT: ID: 0025297 REPORT DATE: 23-AUG-89
ADULT CLINICAL SYSTEM
INTERPRETIVE REPORT

PROFILE VALIDITY

This client's approach to the MMPI-2 was open and cooperative. The resulting MMPI-2 profile is valid and probably a good indication of his present level of personality functioning. This may be viewed as a positive indication of his involvement with the evaluation.

SYMPTOMATIC PATTERNS

The client is exhibiting much somatic distress and may be experiencing a problem with his psychological adjustment. His physical complaints are probably extreme, possibly reflecting a general lack of effectiveness in life. He is probably feeling quite tense and nervous, and may be feeling that he cannot get by without help for his physical problems. He is likely to be reporting a great deal of pain, and feels that others do not understand how sick he is feeling. He may be quite irritable and may become hostile if his symptoms are not given "proper" attention.

Many individuals with this profile have a history of psychophysiological disorders. They tend to overreact to minor problems with physical symptoms. Ulcers and gastrointestinal distress are common. The possibility of actual organic problems, therefore, should be carefully evaluated. Individuals with this profile report a great deal of tension and a depressed mood. They tend to be pessimistic and gloomy in their outlook toward life.

The client seems to have a rather limited range of interests and tends to prefer stereotyped masculine activities over literary and artistic pursuits or introspective experiences. He tends to be somewhat competitive and needs to see himself as masculine. He probably prefers to view women in subservient roles. Interpersonally, he is likely to be intolerant and insensitive, and others may find him rather crude, coarse, or narrow-minded.

In addition, the following description is suggested by the content of this client's responses. He has difficulty managing routine affairs, and the item content he endorsed suggests a poor memory, concentration problems, and an inability to make decisions. He appears to be immobilized and withdrawn and has no energy for life. He views his physical health as failing and reports numerous somatic concerns. He feels that life is no longer worthwhile and that he is losing control of his thought processes.

INTERPERSONAL RELATIONS

He appears to be somewhat passive-dependent in relationships. He may manipulate others through his physical symptoms, and become hostile if sufficient attention is not paid to his complaints. Marital unhappiness is likely to be a factor in his present clinical picture. Many married persons with this profile have difficulties with their marriage. They typically show a diminished interest in sex. Moreover, his moodiness and whining are likely to place additional strains on the marriage.

Figure 16.3. continued

TM page 2
MMPI-2 TM
THE MINNESOTA REPORT: ID: 0025297 REPORT DATE: 23-AUG-89
ADULT CLINICAL SYSTEM
INTERPRETIVE REPORT

He is somewhat shy, with some social concerns and inhibitions. He is a bit hypersensitive about what others think of him and is occasionally concerned over his relationships with others. He appears to be somewhat inhibited in personal relationships and social situations, and may have some difficulty expressing his feelings toward others.

BEHAVIORAL STABILITY

There are likely to be long-standing personality problems predisposing him to develop physical symptoms under stress. His present disorder could reflect, in part, an exaggerated response to environmental stress. Social introversion-extraversion tends to be a very stable personality characteristic over time. His interpersonal shyness is likely to continue to be part of his personality pattern if retested at a later date.

DIAGNOSTIC CONSIDERATIONS

Individuals with this profile type are often seen as neurotic, and may receive a diagnosis of Somatoform Disorder. Actual organic problems such as ulcers and hypertension might be part of the clinical picture. Some individuals with this profile have problems with abuse of pain medication or other prescription drugs.

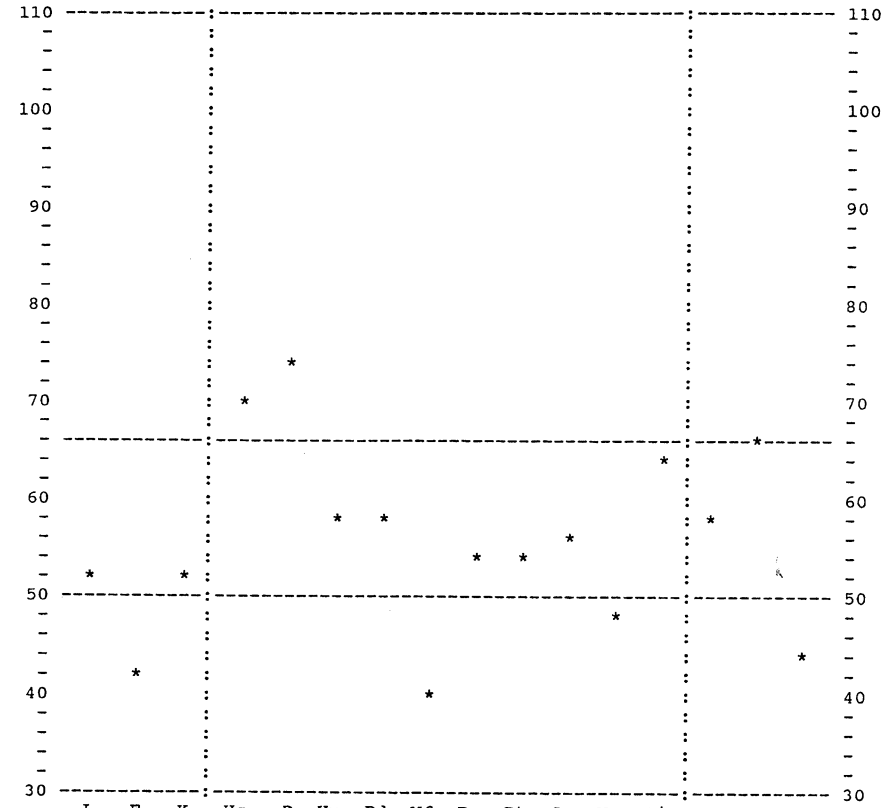
TREATMENT CONSIDERATIONS

He is likely to view his problems as physical and may not readily recognize the psychological factors involved. He tends to somatize his difficulties and to seek medical solutions rather than to deal with them psychologically. He seems to tolerate a high level of psychological conflict and may not be motivated to deal with his problems directly. He is not a strong candidate for insight-oriented psychotherapy. Psychological treatment may progress more rapidly if his symptoms are dealt with through behavior modification techniques. However, with his generally pessimistic attitude and low energy resources, he seems to have little hope of getting better. In any intervention or psychological evaluation program involving occupational adjustment, his negative work attitudes could become an important problem to overcome. He holds a number of attitudes and feelings that could interfere with work adjustment.

NOTE: This MMPI-2 interpretation can serve as a useful source of hypotheses about clients. This report is based on objectively derived scale indexes and scale interpretations that have been developed in diverse groups of patients. The personality descriptions, inferences and recommendations contained herein need to be verified by other sources of clinical information since individual clients may not fully match the prototype. The information in this report should most appropriately be used by a trained, qualified test interpreter. The information contained in this report should be considered confidential.

Figure 16.3. continued

TM page 3
MMPI-2 TM
THE MINNESOTA REPORT: ID: 0025297 RPT DATE: 23-AUG-89
ADULT CLINICAL SYSTEM SEX: Male EDUC: 12
INTERPRETIVE REPORT AGE: 48 MARS: Married
SETTING: Chronic Pain



	L	F	K	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma	Si	A	R	MAC-R
Raw Score:	4	2	16	13	30	24	20	21	11	12	13	16	37	16	22	19
K Corr.				8			6			16	16	3				
T Scr.	52	42	51	70	74	57	57	40	53	53	55	47	63	58	65	44

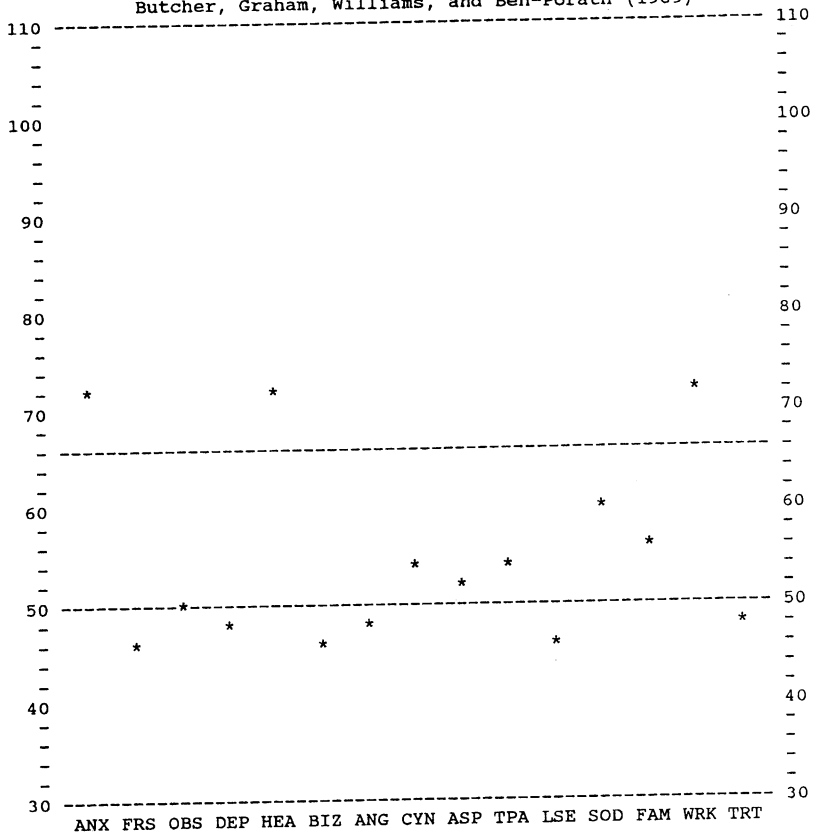
FB (Raw): 2
? Cannot Say (Raw): 0 F-K (Raw): -14 Pain Classification: 0

Welsh Code (new): 21'0-34867/95: LK/F: Percent True : 37
Welsh Code (old): 2"1'4308 7-695/ KL/F?: Percent False: 63
Profile Elev.: 58.3

Figure 16.3. continued

TM page 4
MMPI-2 TM
THE MINNESOTA REPORT: ID: 0025297 RPT DATE: 23-AUG-89
ADULT CLINICAL SYSTEM SEX: Male EDUC: 12
INTERPRETIVE REPORT AGE: 48 MARS: Married
SETTING: Chronic Pain

Content Scales Profile
Butcher, Graham, Williams, and Ben-Porath (1989)



	ANX	FRS	OBS	DEP	HEA	BIZ	ANG	CYN	ASP	TPA	LSE	SOD	FAM	WRK	TRT
Raw Score:	15	2	5	3	15	1	5	13	9	10	2	13	7	19	3
T Score:	72	45	50	48	72	46	48	54	51	53	45	60	55	72	47

Figure 16.3. continued

MMPI-2 TM
THE MINNESOTA REPORT: ID: 0025297 REPORT DATE: 23-AUG-89
ADULT CLINICAL SYSTEM
INTERPRETIVE REPORT

SUPPLEMENTARY SCORE REPORT

	Raw Score	T Score
Ego Strength (Es)	36	47
Dominance (Do)	18	55
Social Responsibility (Re)	16	39
Overcontrolled Hostility (O-H)	10	41
PTSD - Keane (PK)	7	48
PTSD - Schlenger (PS)	20	62
True Response Inconsistency (TRIN)	10	57T
Variable Response Inconsistency (VRIN)	9	65
Depression Subscales (Harris-Lingoes):		
Subjective Depression (D1)	14	69
Psychomotor Retardation (D2)	11	81
Physical Malfunctioning (D3)	6	75
Mental Dullness (D4)	6	67
Brooding (D5)	2	51
Hysteria Subscales (Harris-Lingoes):		
Denial of Social Anxiety (Hy1)	3	45
Need for Affection (Hy2)	4	40
Lassitude-Malaise (Hy3)	6	66
Somatic Complaints (Hy4)	6	67
Inhibition of Aggression (Hy5)	4	55
Psychopathic Deviate Subscales (Harris-Lingoes):		
Familial Discord (Pd1)	1	45
Authority Problems (Pd2)	6	68
Social Imperturbability (Pd3)	4	52
Social Alienation (Pd4)	4	51
Self-Alienation (Pd5)	4	53
Paranoia Subscales (Harris-Lingoes):		
Persecutory Ideas (Pa1)	3	58
Poignancy (Pa2)	3	55
Naivete (Pa3)	4	46

Figure 16.3. continued

MMPI-2 TM
THE MINNESOTA REPORT: ID: 0025297 REPORT DATE: 23-AUG-89
ADULT CLINICAL SYSTEM
INTERPRETIVE REPORT

	Raw Score	T Score
Schizophrenia Subscales (Harris-Lingoes):		
Social Alienation (Sc1)	3	51
Emotional Alienation (Sc2)	0	40
Lack of Ego Mastery, Cognitive (Sc3)	3	60
Lack of Ego Mastery, Conative (Sc4)	5	65
Lack of Ego Mastery, Def. Inhib. (Sc5)	1	47
Bizarre Sensory Experiences (Sc6)	3	55
Hypomania Subscales (Harris-Lingoes):		
Amorality (Ma1)	0	35
Psychomotor Acceleration (Ma2)	6	53
Imperturbability (Ma3)	2	41
Ego Inflation (Ma4)	3	50
Social Introversion Subscales (Ben-Porath, Hostetler, Butcher, & Graham):		
Shyness / Self-Consciousness (Si1)	9	62
Social Avoidance (Si2)	4	54
Alienation--Self and Others (Si3)	5	50

Uniform T scores are used for Hs, D, Hy, Pd, Pa, Pt, Sc, Ma, and the Content Scales; all other MMPI-2 scales use linear T scores.

Figure 16.3. continued

TM
MMPI-2 TM
THE MINNESOTA REPORT: ID: 0025297 REPORT DATE: 23-AUG-89
ADULT CLINICAL SYSTEM
INTERPRETIVE REPORT

CRITICAL ITEMS

The following critical items have been found to have possible significance in analyzing a client's problem situation. Although these items may serve as a source of hypotheses for further investigation, caution should be taken in interpreting individual items because they may have been inadvertently checked.

Acute Anxiety State (Koss-Butcher Critical Items)

10. I am about as able to work as I ever was. (F)
15. I work under a great deal of tension. (T)
208. I hardly ever notice my heart pounding and I am seldom short of breath. (F)
223. I believe I am no more nervous than most others. (F)
301. I feel anxiety about something or someone almost all the time. (T)

Depressed Suicidal Ideation (Koss-Butcher Critical Items)

38. I have had periods of days, weeks, or months when I couldn't take care of things because I couldn't "get going." (T)
233. I have difficulty in starting to do things. (T)

Situational Stress Due to Alcoholism (Koss-Butcher Critical Items)

264. I have used alcohol excessively. (T)
489. I have a drug or alcohol problem. (T)

Mental Confusion (Koss-Butcher Critical Items)

31. I find it hard to keep my mind on a task or job. (T)
299. I cannot keep my mind on one thing. (T)
325. I have more trouble concentrating than others seem to have. (T)

Persecutory Ideas (Koss-Butcher Critical Items)

17. I am sure I get a raw deal from life. (T)
124. I often wonder what hidden reason another person may have for doing something nice for me. (T)

Antisocial Attitude (Lachar-Wrobel Critical Items)

35. Sometimes when I was young I stole things. (T)
105. In school I was sometimes sent to the principal for bad behavior. (T)
254. Most people make friends because friends are likely to be useful to them. (T)
266. I have never been in trouble with the law. (F)

Figure 16.3. continued

TM
MMPI-2 TM
THE MINNESOTA REPORT: ID: 0025297 REPORT DATE: 23-AUG-89
ADULT CLINICAL SYSTEM
INTERPRETIVE REPORT

Somatic Symptoms (Lachar-Wrobel Critical Items)

33. I seldom worry about my health. (F)
53. Parts of my body often have feelings like burning, tingling, crawling, or like "going to sleep." (T)
57. I hardly ever feel pain in the back of my neck. (F)
142. I have never had a fit or convulsion. (F)
159. I have never had a fainting spell. (F)
176. I have very few headaches. (F)
224. I have few or no pains. (F)
247. I have numbness in one or more places on my skin. (T)
295. I have never been paralyzed or had any unusual weakness of any of my muscles. (F)
464. I feel tired a good deal of the time. (T)

Sexual Concern and Deviation (Lachar-Wrobel Critical Items)

121. I have never indulged in any unusual sex practices. (F)

Anxiety and Tension (Lachar-Wrobel Critical Items)

15. I work under a great deal of tension. (T)
17. I am sure I get a raw deal from life. (T)
223. I believe I am no more nervous than most others. (F)
261. I have very few fears compared to my friends. (F)
299. I cannot keep my mind on one thing. (T)
301. I feel anxiety about something or someone almost all the time. (T)
405. I am usually calm and not easily upset. (F)

Deviant Thinking and Experience (Lachar-Wrobel Critical Items)

122. At times my thoughts have raced ahead faster than I could speak them. (T)
427. I have never seen a vision. (F)

Depression and Worry (Lachar-Wrobel Critical Items)

10. I am about as able to work as I ever was. (F)
339. I have sometimes felt that difficulties were piling up so high that I could not overcome them. (T)
415. I worry quite a bit over possible misfortunes. (T)

Figure 16.3. continued

MMPI-2™

MINNESOTA MULTIPHASIC PERSONALITY INVENTORY-2™

By Starke R. Hathaway and J. Charnley McKinley

THE MINNESOTA REPORT™:
PERSONNEL SELECTION SYSTEM
SCREENING REPORT

By James N. Butcher

Client ID: 0025297
Report Date: 22-SEP-89
Age: 48
Sex: Male
Occupation: Other
Education: 12

MINNESOTA MULTIPHASIC PERSONALITY INVENTORY-2
Copyright (c) by THE REGENTS OF THE UNIVERSITY OF MINNESOTA
1942, 1943 (renewed 1970), 1989. This Report 1989.
All Rights Reserved.
Distributed Exclusively by NATIONAL COMPUTER SYSTEMS, INC.
Under License from The University of Minnesota.

"MMPI-2," "Minnesota Multiphasic Personality Inventory-2," and
"The Minnesota Report" are trademarks owned by The University of Minnesota.
Printed in the United States of America.

Figure 16.4. Personnel screening computer-generated report for the individual described in Figures 16.2 and 16.3. Reproduced by permission of the publisher, The University of Minnesota Press.

MMPI-2™
THE MINNESOTA REPORT™:
PERSONNEL SELECTION SYSTEM
SCREENING REPORT

ID: 0025297 REPORT DATE: 22-SEP-89

OPENNESS TO EVALUATION

OVERLY FRANK	QUITE OPEN	ADEQUATE	OVERLY CAUTIOUS	GUARDED	INDETERMINATE
X					

SOCIAL FACILITY

EXCELLENT	GOOD	ADEQUATE	PROBLEMS POSSIBLE	POOR	INDETERMINATE
				X	

ADDICTION POTENTIAL (STANDARD LEVEL)

LOW	NO APPARENT PROBLEM	PROBLEMS POSSIBLE	MODERATE	HIGH	INDETERMINATE
	X				

Figure 16.4. continued

```
                                                                    page 2
     TM
MMPI-2                TM
THE MINNESOTA REPORT:                  ID: 0025297     REPORT DATE: 22-SEP-89
PERSONNEL SELECTION SYSTEM
SCREENING REPORT


                        STRESS TOLERANCE

                                    PROBLEMS
HIGH       GOOD      ADEQUATE       POSSIBLE        LOW        INDETERMINATE
-------------------------------------------------------X----------------------


                        OVERALL ADJUSTMENT

                                    PROBLEMS
EXCELLENT  GOOD      ADEQUATE       POSSIBLE        POOR       INDETERMINATE
-------------------------------------------------------X----------------------

His MMPI-2 responses indicate that his psychological adjustment is likely to
be poor.

This applicant should be evaluated further to determine if he has adjustment
problems.

------------------------------------------------------------------------------
NOTE:  This MMPI-2 report can serve as a useful guide for employment
decisions in which personality adjustment is considered important for
success on the job.  The decision rules on which these classifications are
based were developed through a review of the empirical literature on the
MMPI-2 with "normal-range" individuals (including job applicants) and the
author's practical experience using the test in employment selection.  The
report can assist psychologists and physicians involved in personnel
selection by providing an "outside opinion" about the applicant's
adjustment.  The MMPI-2 should NOT be used as the SOLE means of determining
the applicant's suitability for employment.  The information in this report
should be used by qualified test interpretation specialists ONLY.
------------------------------------------------------------------------------
```

Figure 16.4. continued


```
     TM                                                             page 3
MMPI-2                TM
THE MINNESOTA REPORT:                  ID: 0025297     REPORT DATE: 22-SEP-89
PERSONNEL SELECTION SYSTEM
SCREENING REPORT

                        CONTENT THEMES

The following content themes may serve as a source of hypotheses for
further investigation.  These content themes summarize similar item
responses that appear with greater frequency with this applicant than
with most people.

May be overly sensitive in interpersonal relationships.

May have low self-esteem that interferes with his taking on new tasks.

May show some discomfort in social situations.

May show irresponsible attitudes.

May sometimes disregard rules when it suits him.

May be prone to feeling anxious.

May show low energy or lack of enthusiasm.

May have problems with somatic distress.

May be overly sensitive to criticism.

May have a narrow pattern of interests.

May have negative attitudes or behaviors that would interfere with work
adjustment.
```

Figure 16.4. continued

most patients" (Erdman, Greist, Klein, Jefferson, & Getto, 1981, p. 394). In general, concerns about the impersonality of such interactions have not proved justified. Reports show that patients usually enjoy the process and have even revealed more sensitive data to a computer than to a human interviewer (Erdman, Klein, & Greist, 1985; Greist & Klein, 1980, 1981; Johnson & Williams, 1980; Lucas, Mullins, Luna, & McInroy, 1977). However, the latter point illustrates that computer-administered versions of tests cannot necessarily be assumed to be parallel forms of paper-and-pencil versions. Moreland (1987) notes that "rather subtle differences between on-line and conventional test administrations can produce large differences in test scores produced by the two modes" (p. 33). For example, providing a clear "Cannot Say" option along with the "True" and "False" response options on the MMPI or CPI changes scores by producing more omitted items than on booklet forms where the "Cannot Say" option is not as obvious (Biskin & Kolotkin, 1977; Lushene, O'Neil, & Dunn, 1974; Scissons, 1976).

Based on his review, Moreland concludes that if care is taken to make the computerized administrative format as similar as possible to the paper-and-pencil form, "non-equivalence is typically small enough to be of no practical consequence, if present at all" (p. 34). However, further research is needed to delineate the important parameters determining test equivalence, and to address individual differences in responses to computerized test administration.

Applications such as these and the many others described in the contemporary literature illustrate automated technology's potential for streamlining traditional objective personality assessment procedures. Computers can increase the accuracy and reliability of standardized administration, data collection, data storage, and application of objective decision rules. However, online technology also offers the possibility of more innovative uses than simply duplicating exist-

ing manual procedures. Some new uses include programming computers to mimic the human judgmental process, such as in the narrative interpretive schemes described above. The next sections will describe applications using the computer's capacity for dynamic, adaptive systems of analysis to improve human decision making.

Dynamic, Adaptive Assessment Strategies

Although Holtzman (1960) claimed that the clinician's uniquely human adaptive, creative, and intuitive judgment could never be reproduced by a machine, Kleinmuntz and McLean (1968) listed flexibility as one of the three main advantages of a computer—along with speed and objectivity. According to them, "The ability of the computer to alter its activity as the result of the environment in which it is working is its greatest strength" (p. 75). In the case of an online test administration or interview, the computer can be programmed to choose one of several alternative content areas as a function of an individual's prior responses, much as a human interviewer would. Such a capability can be very useful in psychiatric screening, both in tailoring specific assessment devices to maximize information about particular patients and in adapting a standardized interview sequence to the characteristics and needs of the individual subject. In addition, the flexibility and perfect memory of the computer make it possible for it to evaluate its own performance by specified criteria and to modify subsequent decisions and actions accordingly.

Branched Interviews

Most of the clinical work utilizing flexible assessment strategies has involved adaptive interviews. The growing utilization of large automated patient data systems requires collection of great quantities of information on each individual, but administration of the same lengthy interview to every subject can often be a wasteful and inefficient procedure. Certain responses make other questions inappropriate or, on the other hand, may indicate the need for branching to more detailed questioning in specific topic areas. For example, asking an unemployed patient a series of questions relating to job performance would be an inefficient use of patient and interviewer time, and knowing that a patient is married may indicate the need to ask further questions about marital adjustment.

Several automated interview systems have been designed to deal with these problems. Stillman, Roth, Colby, and Rosenbaum (1969) described one of the first online psychiatric inventories, the Computer-Assisted Special Enquirer (CASE) developed at Stanford. The clinician entered interview questions and branching options into a terminal, using both multiple choice and narrative format items. This interview network was then presented to patients via an online terminal. Any particular patient received only about 50% of the total item pool due to the branching options. A computer-generated report indicated to the clinician which items were presented and what the responses to these were.

Erdman, Klein, and Greist (1985; Greist & Klein, 1980) at the University of Wisconsin have described adaptive administration of their computerized Social Adjustment Interview, which collects patient history information and evaluates current role functioning. Detailed questions later in the interview are conditional upon earlier answers to a standard set of demographic items and the patient's choice of "two areas that are most involved in your problems right now" (1980, p. 165). Although the interview as a whole consists of more than 650 questions, the average single patient is likely to be presented with about 150 of these items. Erdman et al. reported excellent agreement between clinical evaluations and results of the computerized interview. Similarly, the computer-based Psychiatric Assessment Unit developed at the Veterans Administration Hospital in Salt Lake City (Johnson & Williams, 1980) has been reported to result in faster, cheaper, more accurate and consistent reports than the hospital's traditional assessment methods (Klingler, Miller, Johnson, & Williams, 1976, 1977).

Adaptive Test Administration

Adaptive administration of objective personality tests is not as well researched as flexible interview strategies. Although theoretically the branching ability of the computer could make it possible to produce personality tests "tailored to each individual's needs and abilities" (Patience, 1977), thus far the application of adaptive testing strategies to achievement and ability testing (English, Reckase, & Patience, 1977; Vale & Weiss, 1975; Weiss, 1973) still greatly surpasses their use in personality assessment. Many foresee personality assessment relying more and more in the future on individualized, tailored tests (e.g., Bartram & Bayliss, 1984; Ben-Porath & Butcher, 1986; Butcher, 1986; Butcher, Keller, & Bacon, 1985).

Research investigating the use of adaptive testing strategies in personality assessment falls roughly into four categories, based on the method used: (a) the prediction of full scores on the conventional paper-and-pencil form of the test, (b) the adaptive typological approach, (c) the countdown strategy, and (d) methods based on item response theory (IRT, also known as latent trait theory). In all four of these approaches, the majority of research has focused on examining the merits of altering extant personality assessment instruments (such as the MMPI) by applying adaptive testing methods.

Kleinmuntz and McLean (1968), pioneers in the use of adaptive testing in personality assessment, presented a model for a branched, interactive, computerized MMPI short form. Their program presented the subject with a subset of five items from each of the 15 scales to be estimated, on the basis of which a T-score was computed for each scale. Additional items were administered only for those scales that could not be clearly classified as "normal" or "abnormal" based on these T-scores, and the test ended when all scales could be so classified. Kleinmuntz and McLean's system cut the MMPI administration time drastically, but the correlations of their scale scores with the long form scores were not impressive.

More recently, Sanders (1985) explored the feasibility of developing individualized short forms of the MMPI. He used a two-stage approach, as did Kleinmuntz and McLean. Initially five items for each of 12 MMPI scales (the *Mf* scale was not scored) were administered via computer. Additional items were administered for each scale and entered into a stepwise multiple regression equation. When the predicted full-scale score was not altered significantly by the administration of additional items, item sampling for that scale was terminated and the computer went on to the next scale. The items that were administered for each scale were those that had the highest item-scale correlations. Sanders was successful in obtaining high correlations between predicted MMPI scale scores and full-scale scores, but the correlations between predicted and actual code types were not reported.

Clavelle and Butcher (1977) utilized an adaptive typological approach to predict MMPI code type membership, rather than estimating long form scale scores as Kleinmuntz and McLean (1968) and Sanders (1985) had done. Hoffman and Butcher (1975) have pointed out that even though good prediction of long form scale elevations can be achieved by several popular MMPI short forms, the congruence between code type classification based on the two forms is generally too low to be useful for practical clinical decision making. They suggested that instead of attempting to predict long form scale elevations, "a more fruitful clinical approach might be to use multivariate techniques, such as the discriminant function, to select a small subset of items for direct prediction of selected criteria or for specific decision problems" (p. 38).

Clavelle and Butcher implemented this suggestion by examining item responses for a large group of outpatients whose MMPI profiles fell into one of nine selected code type groups. Sixty-nine items which showed significant differences in endorsement frequency between these groups were used as predictor variables in a stepwise multiple discriminant function analysis with code type as the criterion. Accuracy of classification was assessed after administration of each block of 10 items. Results showed that classification accuracy tended to reach an asymptotic level after administration of a relatively small number of items, and in addition, certain blocks of code types tended to form higher order groups that were retained or excluded together. Clavelle and Butcher concluded that

> rather than continue with the administration of items of increasingly dubious discriminant value, it seems that a more efficient strategy would be to branch out to another set of items capable of providing the most additional information pertinent to a differentiation among the remaining code types. (p. 857)

Another approach which can also be considered typological, the countdown method, was described by Butcher, Keller, and Bacon (1985). When this strategy is applied to MMPI administration, it can be used to classify individuals more quickly as "normal" or "abnormal," i.e. whether their T-score on a particular scale is either less than 70 or is 70 or greater. Item administration for each scale is terminated when the number of unendorsed items equals the total number of items on the scale minus the cutoff (the number of endorsed items which correspond to a T-score of 70) plus one, or when the items answered in the keyed direction add up to the cutoff. For example, if a scale is 30 items long, and 20 endorsed items are required to obtain a T-score of 70 (the cutoff), as soon as an individual fails to endorse 11 items, administration of that scale can be terminated because there is no possibility of the subject exceeding a T-score of 70. That person has been classified as "normal" with complete accuracy as judged by full form scores.

Recently, Slutske, Ben-Porath, and Butcher (1988) described an implementation of the countdown

method. They simulated an adaptive administration of the MMPI using data already collected from subjects who had completed the full paper-and-pencil version. In addition, they looked at the efficiency of ordering the items so that the least frequently endorsed items were presented first. They found that, particularly with a personnel selection sample, using the countdown method and administering items in ascending order of endorsement frequency resulted in a substantial savings in the number of items required. With a psychiatric sample, the most efficient way to administer items was to present the items most frequently endorsed by a normative sample first. Clearly, different approaches are required when dealing with individuals in which normal-range profiles are expected (personnel selection) versus when elevated profiles are more common (psychiatric sample). Slutske et al. suggest that the countdown strategy may be very useful for screening purposes where length of test is an issue.

A more psychometrically sophisticated adaptive testing technique, already extensively used in the areas of achievement and ability testing, is based on item-response theory (IRT). Briefly, according to Weiss (1985), IRT-based adaptive testing

> selects items that provide maximum levels of item information at an individual's currently estimated trait level. In addition, IRT-based methods of scoring tests permit estimation of individuals' trait levels based on their responses to one or more items. As a consequence, an item can be administered and an estimate can be made of the individual's level on the trait. After the administration of an item and estimation of trait, the new trait level is used to select the next item to be administered to that examinee. (p. 783)

The benefits of IRT-based adaptive testing are that it can decrease testing time by as much as 50% while maintaining or even improving on the precision of measurement obtained via conventional tests. For a concise review of this field, see Weiss (1985) or Weiss and Vale (1987).

The application of IRT-based adaptive testing to personality assessment is still a new and largely unexplored area. One of the first attempts to apply the IRT model to personality variables was a doctoral dissertation by Sapinkopf (1978). He compared the adaptive administration of several scales of the CPI to both the conventional paper-and-pencil format and to a nonadaptive computer-administered version. Sapinkopf found the adaptive strategy to be much more efficient than the other two methods in both the number of items and the testing time required, although the reliabilities of two of the three adaptively administered CPI scales were lower than the reliabilities of the conventionally administered scales.

Carter (1982), again in a doctoral dissertation, investigated using a different IRT model to adaptively administer the MMPI. She obtained rather low correlations between the profiles obtained with the tailored and full-scale versions. In fact, her adaptively developed short form did not perform as well as two commonly used conventional short forms. Various reasons for the low correlations were suggested, including the possibility that using the more complicated IRT model employed by Sapinkopf (1977) might have resulted in more accurate estimates of full-scale scores with the same reduced number of items.

Carter and Wilkinson (1984) went on to identify those items on the MMPI which are "bad" according to latent trait models. Items which fit the model poorly have low or even negative discriminations: individuals low on the dimension being measured tend to endorse the item, while those high on the dimension do not, a result opposite to what is desired or expected. It was found that each of the basic MMPI scales had at least a few such items. It is interesting that the poorly fitting items according to their IRT analysis were the so-called subtle items originally identified by Weiner (1956), which some believe were included on the MMPI scales through chance correlations. Carter and Wilkinson concluded that the MMPI might be suitable for IRT-based adaptive administration if a two-parameter model (which takes item discriminations into account) is used.

Contrary to Carter and Wilkinson's conclusions, some experts believe that the application of IRT-based adaptive testing models to instruments such as the MMPI and CPI is inappropriate because these tests are empirically keyed and thus their scales are not unidimensional (cf. Ben-Porath and Butcher, 1986). The use of such tests results in the violation of some of the major assumptions of IRT, especially trait unidimensionality. Factorially derived personality assessment scales are much more appropriate for IRT methods, as these types of instruments are more likely to contain homogeneous, unidimensional scales.

It has recently been demonstrated (Reise & Waller, in press; Waller & Reise, in press) that adaptive testing using IRT works well for personality scales which meet the assumptions of the IRT model. The researchers used a computerized adaptive version of Tellegen's Multidimensional Personality Questionnaire (MPQ) (Tellegen, 1982), a factor analytically developed personality assessment instrument made up of 11 scales tapping various aspects of normal personality functioning. They were able to administer an individually tailored set of items chosen on the basis of the respondent's unique sequence of item endorsements and the difficulties and discriminations of the endorsed items. Reise and Waller found that it was possible to attain accurate trait estimates while maintaining an average item savings in excess of 40%.

There are several considerations in evaluating adaptively administered versions of objective personality tests. Just as with any computer-administered test, the equivalence of the adaptive version to the traditional paper-and-pencil version must be established empirically. The different response characteristics required may alter the nature of the instrument, and may make the norms developed on the traditional form inapplicable to the computer-administered version.

There are also considerations unique to adaptive versions of preexisting personality inventories. The adaptive administration of traditional instruments involves the rearrangement of the items within the inventory. Researchers in the area of achievement testing (Kingston & Dorans, 1984) have found context, or item-location effects, to be important on such tests as the Graduate Record Examination (GRE). Thus far, there are no studies investigating the importance of context for personality items.

Both of these considerations ultimately concern the validity of the modified versions of traditional personality assessment instruments. Additionally, many of the studies reviewed here were simulation studies: response data obtained from the conventional administration of personality assessment instruments were used to investigate the potential merits of adaptive versions of these tests. Until the adaptive versions are interactively administered to real subjects sitting at computer terminals, these studies must be considered preliminary. In addition, although attempts at modifying existing personality assessment instruments through adaptive testing strategies have been encouraging, it is likely that more benefit will be achieved from the development of new tests specifically designed for adaptive administration.

Development of New Assessment Strategies.

Johnson, Giannetti, and Williams (1979) noted that automation of standard clinical techniques does not take full advantage of the computer's capabilities and data bank. Accordingly, they attempted to construct a new inventory specifically designed for the computer. They took all the items in the data base of the Salt Lake City Veterans' Administration Hospital's computerized Psychological Assessment Unit (approximately 2500 questions based on a large battery of standardized tests) and had expert clinicians categorize them rationally into content areas. They also took suggestions for items to be added for underrepresented content areas, and deleted redundant or overrepresented questions. The full question list was then administered to normal and psychiatric samples, and factor analysis was used to extract the empirical factors within each content domain. Only the items with the highest loadings on these factors were retained in the item pool. (This approach is similar to the multistage, multimethod strategy used in developing new content scales for the revised MMPI.) Johnson's group also added demographic and historical items at the beginning of the test as a basis for branching through the rest of the item pool, and empirically derived a dissimulation index to check on response validity. The preliminary version of the Psychological Systems Questionnaire (PSQ) consisted of 24 multiple choice demographic and branching items, 18 true-false dissimulation index items, and 729 true-false items which were scored on several factor analytically derived scales.

Several aspects of the PSQ's functioning illustrate the tremendous future potential of interactive computerized personality assessment. First, the demographic items at the beginning save time by allowing elimination of later items that are irrelevant or inappropriate. Second, the computer's ongoing computational ability allows it continually to calculate the individual's current standing on each scale throughout the entire test administration and to respond appropriately. For example:

> When a respondent is found to have scored high on a particular scale, he/she is asked for comments . . . the computer might say, "Mr. Jones, your test responses seem to indicate that you are very depressed. Would you care to make any comments about this?" These comments are then taken free-form from the typewriter keyboard. (Johnson, Giannetti, & Williams, 1979, p. 259)

This continuous scale-scoring ability may prove to be particularly useful in evaluating the validity of a patient's responses to an interactive self-report questionnaire. Evaluation of validity has always been a difficult problem in scoring self-report personality tests. The PSQ program was designed to print out cautionary reminders to the subject if the dissimulation index was getting high, and to terminate the test prematurely if the response pattern seemed completely invalid. Stout (1981) reports an even more

sophisticated use of computer capabilities in ongoing evaluation of response validity. He suggests incorporation of ancillary data such as response latencies and key pressure into the computer's decision scheme. If certain latency patterns are found to be reliable indicators of invalidity or other unusual testing events, the computer could respond to these patterns appropriately. Stout lists examples, such as sudden high-speed responding as an indicator of randomness, or progressively longer response latencies as an indicator of fatigue. The computer could respond to such patterns either passively, by marking potentially invalid responses, or actively, by recommending a rest break, encouraging better cooperation and then readministering items, inserting extra items to check on the validity of earlier ones, switching to an alternate form more appropriate for the subject's capabilities, or signaling to a human examiner that something is amiss in the testing situation. Even computerized voice analysis might eventually be employed as a measure of test-taking attitude (Ben-Porath & Butcher, 1986).

Finally, one of the most important advantages of computerized testing systems is the computer's ability to utilize empirical data to modify its decision rules. If the computer is provided with information on the accuracy and utility of its decisions, it could be programmed to evaluate continually the relative contribution of particular items, actuarial formulas, etc., and drop or add variables and decision rules as appropriate. Computer systems will be ideal testing grounds for new measures and assessment strategies (Ben-Porath & Butcher, 1986; Butcher, Keller, & Bacon, 1985; Giannetti, Klingler, Johnson, & Williams, 1976).

CONCLUSION

D. J. Smail (1971) criticized what he saw as psychology's infatuation with technology, commenting that

> by becoming dazzled by statistical sophistication and technological efficiency, he [the psychologist] may forget to question the basis of his science and the ultimate purpose of his actions. In consequence he stands in danger of building around him an artificial world which, although possibly beautifully efficient within its own boundaries and elegant in its design, may bear no relation to anything that matters. (p. 173)

This review has attempted to show that, despite Smail's claims, personality assessment technology can and does have relevance to many things that matter. We pointed out at the beginning of this chapter that the value of technology lies in its utility as a tool to achieve valued goals. The extensive research and clinical literature on many objective personality tests attests to the practical value of these tools for contemporary psychological applications. Future developments, including new test construction methods and the increasing use of computer technology, will have to face the same empirical tests as have made or broken past assessment strategies. However, we believe that objective tests and their logical extension—automated methods—will play a major, successful role in the future of personality assessment technology.

REFERENCES

Adair, L. F. (1978). Review of MMPI computer services. In O. Buros (Ed.), *Eighth mental measurements yearbook*. Highland Park, NJ: Gryphon.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychological Association. (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC: Author.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Barron, F. (1953). An ego-strength scale which predicts response to psychotherapy. *Journal of Consulting Psychology, 17,* 327–333.

Bartram, D., & Bayliss, R. (1984). Automated testing: Past, present, and future. *Journal of Occupational Psychology, 57,* 221–237.

Baucom, D. H. (1985). Review of California Psychological Inventory. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 250–252). Lincoln: University of Nebraska Press.

Ben-Porath, Y. S., & Butcher, J. N. (1986). Computers in personality assessment: A brief past, an ebullient present, and an expanding future. *Computers in Human Behavior, 2,* 167–182.

Ben-Porath, Y. S., & Butcher, J. N. (1988, March). *Exploratory analyses of rewritten MMPI items*. Paper presented at the 23rd Annual Symposium on Recent Developments in the Use of the MMPI, St. Petersburg, Florida.

Ben-Porath, Y. S., & Butcher, J. N. (In press). The historical development of personality assessment. In C. E. Walker (Ed.), *Clinical psychology: Historical and research roots*. New York: Plenum Press.

Biskin, B. H., and Kolotkin, R. C. (1977). Effects of computerized administration on scores on the Minnesota Multiphasic Personality Inventory. *Applied Psychological Measurement, 1,* 543–549.

Block, J. (1965). *The challenge of response sets: Unconfounding meaning, acquiescence, and social desirability in the MMPI*. New York: Appleton-Century-Crofts.

Bolton, B. (1985). Review of the Adult personality Inventory. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 55–56). Lincoln: University of Nebraska Press.

Burisch, M. (1984). Approaches to personality inventory construction: A comparison of merits. *American Psychologist, 39,* 214–227.

Butcher, J. N. (1978). Present status of computerized MMPI reporting services. In O. Buros (Ed.), *Eighth mental measurements yearbook* (pp. 942–945). Highland Park, NJ: Gryphon.

Butcher, J. N. (1985). Review of Sixteen personality Factor Questionnaire. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 1391–1392). Lincoln: University of Nebraska Press.

Butcher, J. N. (1986, August). *Future directions in computerized psychological assessment procedures: The psychological clinic in the year 2001*. Paper presented at the 94th Annual Meeting of the American Psychological Association, Washington, DC

Butcher, J. N. (1987). *Computerized psychological assessment: A practitioner's guide*. New York: Basic Books.

Butcher, J. N. (1989a). *User's Guide for the MMPI-2 Minnesota Report: Adult Clinical System*. Minneapolis, MN: University of Minnesota, distributed by National Computer Systems.

Butcher, J. N. (1989b). *User's Guide for the MMPI-2 Minnesota Report: Personnel Selection System*. Minneapolis, MN: University of Minnesota, distributed by National Computer Systems.

Butcher, J. N., Graham, J. R., Dahlstrom, W. G., Tellegen, A. M., & Kaemmer, B. (1989). *MMPI-2 Manual for Administration and Scoring*. Minneapolis: University of Minnesota Press.

Butcher, J. N., Graham, J. R., Williams, C. L., & Ben-Porath, Y. S. (1989). *Development and Use of the MMPI-2 Content Scales*. Minneapolis, MN: University of Minnesota Press.

Butcher, J. N., & Keller, L. S. (1984). Objective personality assessment. In G. Goldstein & M. Hersen (Eds.), *Handbook of psychological assessment* (pp. 307–331). New York: Pergamon Press.

Butcher, J. N., Keller, L. S., & Bacon, S. F. (1985). Current developments and future directions in computerized personality assessment. *Journal of Consulting and Clinical Psychology, 53,* 803–815.

Butcher, J. N., & Owen, P. L. (1978). Objective personality inventories: Recent research and some contemporary issues. In B. Wolman (Ed.), *Handbook of clinical diagnosis of mental disorders* (pp. 475–545). New York: Plenum Press.

Butcher, J. N., & Pancheri, P. (1976). *A handbook of cross-national MMPI research*. Minneapolis: University of Minnesota Press.

Carter, J. E. (1982). *A computerized adaptive version of the MMPI based on the Rasch model*. Unpublished doctoral dissertation, University of Chicago, Chicago Circle.

Carter, J. E., & Wilkinson, L. (1984). A latent trait analysis of the MMPI. *Multivariate Behavioral Research, 19,* 385–407.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire (16PF)*. Champaign, IL: Institute for Personality and Ability Testing.

Clavelle, P. R., & Butcher, J. N. (1977). An adaptive typological approach to psychiatric screening. *Journal of Consulting and Clinical Psychology, 45,* 851–859.

Colligan, R. C., Osborne, D., Swenson, W. M., & Offord, K. P. (1984). *The MMPI: A contemporary normative study*. New York: Praeger.

Comrey, A. L. (1957). Factors in the items of the MMPI. *American Psychologist, 12,* 437.

Comrey, A. L. (1970). *EDITS manual for the Comrey Personality Scales*. San Diego, CA: Educational and Industrial Testing Service.

Costa, P. T., Zonderman, A. B., McCrae, R. R., & Williams, R. B. (1985). Content and comprehensiveness in the MMPI: An item factor analysis in a normal adult sample. *Journal of Personality and Social Psychology, 48,* 925–933.

Dahlstrom, W. G., & Dahlstrom, L. (Eds.). (1980). *Basic readings on the MMPI: A new selection on personality measurement*. Minneapolis: University of Minnesota Press.

Dahlstrom, W. G., Lachar, D., & Dahlstrom, L. E. (1986). *MMPI patterns of American minorities*. Minneapolis: University of Minnesota Press.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook Vol. 1: Clinical interpretation* (rev. ed.). Minneapolis: University of Minnesota Press.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1975). *An MMPI handbook Vol. 2: Research*

applications (rev. ed.). Minneapolis: University of Minnesota Press.

Davies, J. D. (1955). *Phrenology fad and science: A 19th century American crusade*. New Haven, CT: Yale University Press.

Diehl, L. A. (1977). The relationship between demographic factors, MMPI scores and the social readjustment rating scale. *Dissertation Abstracts International, 38* (5–B), 2360.

Drake, L. E. (1946). A social I.E. scale for the MMPI. *Journal of Applied Psychology, 30,* 51–54.

Duckworth, J. C., & Anderson, W. P. (1986). *MMPI interpretation manual for counselors and clinicians* (3rd ed.). Muncie, IN: Accelerated Development.

English, R. A., Reckase, M. D., & Patience, W. M. (1977). Application of tailored testing to achievement measurement. *Behavior Research Methods and Instrumentation, 9,* 158–161.

Erdman, H. P., Greist, J. H., Klein, M. H., Jefferson, J. W., & Getto, C. (1981). The computer psychiatrist: How far have we come? Where are we heading? How far dare we go? *Behavior Research Methods and Instrumentation, 13,* 393–398.

Erdman, H. P., Klein, M. H., & Greist, J. H. (1985). Direct patient computer interviewing. *Journal of Consulting and Clinical Psychology, 53,* 760–773.

Evans, R. G. (1983). Who published what about whom? *Journal of Personality Assessment, 47,* 339–344.

Exner, J. E. (1987). Computer assistance in Rorschach interpretation. In J. N. Butcher (Ed.), *Computerized psychological assessment: A practitioner's guide* (pp. 218–235). New York: Basic Books.

Eyde, L. D. (1985). Review of the Minnesota Report: Personnel Selection System for the MMPI. In J. V. Mitchell (Ed)., *The ninth mental measurements yearbook* (pp. 1005–1008). Lincoln: University of Nebraska Press.

Eyde, L. D., Kowal, D. M., & Fishburne, F. J. (1987, September). *Clinical implications of validity research on computer-based test interpretations of the MMPI*. Paper presented at the Symposium on Practical Test User Problems Facing Psychologists in Private Practice, New York, New York.

Eysenck, H. J. (1985). Review of California Psychological Inventory. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 252–253). Lincoln: University of Nebraska Press.

Fowler, R. D. (1985). Landmarks in computer-assisted psychological testing. *Journal of Consulting and Clinical Psychology, 53,* 748–759.

Fowler, R. D. (1987). Developing a computer-based test interpretation system. In J. N. Butcher (Ed.), *Computerized psychological assessment: A practitioner's guide* (pp. 50–63). New York: Basic Books.

Fowler, R. D., & Butcher, J. N. (1986). Critique of Matarazzo's views on computerized testing: All sigma and no meaning. *American Psychologist, 41,* 94–95.

Galton, F. (1884). Measurement of character. *Fortnightly Review, 42,* 179–185.

Giannetti, R. A., Klingler, D. E., Johnson, J. H., & Williams, T. A. (1976). The potential for dynamic assessment systems using on-line computer technology. *Behavior Research Methods and Instrumentation, 8,* 101–103.

Gilberstadt, H., & Duker, J. (1965). *A handbook for clinical and actuarial MMPI interpretation*. Philadelphia: Saunders.

Goldberg, L. R. (1965). Diagnosticians vs. diagnostic signs: The diagnosis of psychosis vs. neurosis from the MMPI. *Psychological Monographs, 79,* (Entire No. 602)

Goldberg, L. R. (1972). Parameters of personality inventory construction and utilization: A comparison of prediction strategies and tactics. *Multivariate Behavioral Research Monographs,* No. 72–2.

Gough, H. G. (1975). *California Psychological Inventory* (revised manual). Palo Alto, CA: Consulting Psychologists Press.

Graham, J. R. (1987). *The MMPI: A practical guide* (2nd ed.). New York: Oxford University Press.

Graham, J. R., & Butcher, J. N. (1988, March). *Differentiating schizophrenic and major affective disordered inpatients with the revised form of the MMPI*. Paper presented at the 23rd Annual Symposium on Recent Developments in the Use of the MMPI, St. Petersburg, Florida.

Graham, J. R., Schroeder, H. E., & Lilly, R. S. (1971). Factor analysis of items on the Social Introversion and Masculinity-Femininity scales of the MMPI. *Journal of Clinical Psychology, 27,* 367–370.

Grayson, H. M. (1951, June). *Psychological admissions testing program and manual*. Los Angeles: Veterans Administration Center, Neuropsychiatric Hospital.

Green, C. J. (1982). The diagnostic accuracy and utility of MMPI and MCMI computer interpretive reports. *Journal of Personality Assessment, 46,* 359–365.

Greene, R. L. (1980). *The MMPI: An interpretive manual*. New York: Grune & Stratton.

Greer, S. E. (1984). A review of the Millon Clinical Multiaxial Inventory. *Journal of Counseling and Development, 63,* 262–263.

Greist, J. H., & Klein, M. H. (1980). Computer programs for patients, clinicians, and researchers in psychiatry. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 161–181). Norwood, NJ: Ablex.

Greist, J. H., & Klein, M. H. (1981). Computers in psychiatry. In S. Arieti (Ed.), *American handbook of psychiatry* (2nd ed.) (Vol. 7, pp. 750–77). New York: Basic Books.

Guilford, J. P., & Zimmerman, W. S. (1956). Fourteen dimensions of temperament. *Psychological Monographs, 70* (Entire no. 417).

Guthrie, G. (1985). Review of Clinical Analysis Questionnaire. In J. V. Mitchell, Ed., *The ninth mental measurements yearbook* (pp. 340–341). Lincoln: University of Nebraska Press.

Halbower, C. C. (1955). *A comparison of actuarial versus clinical prediction to classes discriminated by MMPI*. Unpublished doctoral dissertation, University of Minnesota.

Harris, R. E., & Lingoes, J. C. (1968). *Subscales for the Minnesota Multiphasic Personality Inventory*. Unpublished manuscript. San Francisco: Langley Porter Clinic.

Hartshorne, H., & May, M. A. (1928). *Studies in the nature of character: Vol. 1. Studies in deceit*. New York: Macmillan.

Hase, H. D., & Goldberg, L. R. (1967). Comparative validity of different strategies of constructing personality inventory scales. *Psychological Bulletin, 67,* 231–248.

Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology, 10,* 249–254.

Henrichs, T. (1964). Objective configural rules for discriminating MMPI profiles in a psychiatric population. *Journal of Clinical Psychology, 20,* 157–159.

Henrichs, T. (1966). A note on the extension of MMPI configural rules. *Journal of Clinical Psychology, 22,* 51–52.

Hess, A. K. (1985). Review of Millon Clinical Multiaxial Inventory. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 984–986). Lincoln: University of Nebraska Press.

Hoffman, N. G., & Butcher, J. N. (1975). Clinical limitations of three Minnesota Multiphasic Personality Inventory short forms. *Journal of Consulting and Clinical Psychology, 43,* 32–39.

Holtzman, W. H. (1960). Can the computer supplant the clinician? *Journal of Clinical Psychology, 16,* 119–122.

Hsu, L. M. (1984). MMPI T-scores: Linear versus normalized. *Journal of Consulting and Clinical Psychology, 52,* 821–823.

Jackson, D. N. (1971). The dynamics of structured personality tests: 1971. *Psychological Review, 78,* 229–248.

Jackson, D. N. (1972). *Differential Personality Inventory*. London, Ontario: Author.

Jackson, D. N. (1984). *Personality Research Form manual*. Port Huron, MI: Research Psychologists Press.

Johnson, J. H., Butcher, J. N., Null, C., & Johnson, K. N. (1984). Replicated item level factor analysis of the full MMPI. *Journal of Personality and Social Psychology, 47,* 105–114.

Johnson, J. H., Giannetti, R. A., & Williams, T. A. (1976). Computers in mental health care delivery: A review of the evolution toward interventionally relevant on-line processing. *Behavior Research Methods and Instrumentation, 8,* 83–91.

Johnson, J. H., Giannetti, R. A., & Williams, T. A. (1979). Psychological Systems Questionnaire: An objective personality test designed for on-line computer presentation, scoring, and interpretation. *Behavior Research Methods and Instrumentation, 11,* 257–260.

Johnson, J. H., & Williams, T. A. (1980). Using on-line computer technology to improve service response and decision-making effectiveness in a mental health admitting system. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 237–249). Norwood, NJ: Ablex.

Karson, S., & O'Dell, J. W. (1975). A new automated interpretation system for the 16PF. *Journal of Personality Assessment, 39,* 256–260.

Karson, S., & O'Dell, J. W. (1987). Computer-based interpretation of the 16PF: The Karson Clinical Report in contemporary practice. In J. N. Butcher (Ed.), *Computerized psychological assessment: A practitioner's guide* (pp. 198–217). New York: Basic Books.

Keller, L. S., & Butcher, J. N. (In press). *Assessment*

of chronic pain patients with the MMPI-2. Minneapolis, MN: University of Minnesota Press.

Kingston, N. M., & Dorans, N. J. (1984). Item location effects and their implications for IRT equating and adaptive testing. *Applied Psychological Measurement, 8,* 147–154.

Kleinmuntz, B. (1982). *Personality and psychological assessment*. New York: St. Martin's Press.

Kleinmuntz, B., & McLean, R. S. (1968). Computers in behavioral science: Diagnostic interviewing by digital computer. *Behavioral Science, 13,* 75–80.

Klingler, D. E., Miller, D. A., Johnson, J. H., & Williams, T. A. (1976). Strategies in the evaluation of an on-line computer-assisted unit for intake assessment of mental health patients. *Behavior Research Methods and Instrumentation, 8,* 95–100.

Klingler, D. E., Miller, D. A., Johnson, J. H., & Williams, T. A. (1977). Process evaluation of an on-line computer-assisted unit for intake assessment of mental health patients. *Behavior Research Methods and Instrumentation, 9,* 110–116.

Koss, M. P. (1980). Assessing psychological emergencies with the MMPI. In J. N. Butcher, G. Dahlstrom, M. Gunther, & W. Schofield (Eds.), *Clinical Notes on the MMPI*. Roche Psychiatric Service Institute Monograph Series. Nutley, NJ: Hoffman–LaRoche.

Koss, M. P., & Butcher, J. N. (1973). A comparison of psychiatric patients' self-report with other sources of clinical information. *Journal of Research in Personality, 7,* 225–236.

Koss, M. P., Butcher, J. N., & Hoffman, N. G. (1976). The MMPI critical items: How well do they work? *Journal of Consulting and Clinical Psychology, 44,* 921–928.

Krug, S. E. (1980). *Clinical Analysis Questionnaire Manual*. Champaign, IL: Institute for Personality and Ability Testing.

Lachar, D. (1974). *The MMPI: Clinical assessment and automated interpretation*. Los Angeles: Western Psychological Services.

Lachar, D., & Wrobel, T. A. (1979). Validation of clinicians' hunches: Construction of a new MMPI critical item set. *Journal of Consulting and Clinical Psychology, 47,* 277–284.

Lewandowski, D., & Graham, J. R. (1972). Empirical correlates of frequently occurring two-point MMPI code types: A replicated study. *Journal of Consulting and Clinical Psychology, 39,* 467–472.

Lucas, R. W., Mullins, P. J., Luna, C. B., & McInroy, D. C. (1977). Psychiatrists and a computer as interrogators of patients with alcohol-related illnesses: A comparison. *British Journal of Psychiatry, 131,* 160–167.

Lushene, R. E., O'Neil, H. F., & Dunn, T. (1974). Equivalent validity of a completely computerized MMPI. *Journal of Personality Assessment, 38,* 353–361.

MacAndrew, C. (1965). The differentiation of male alcoholic outpatients from nonalcoholic psychiatric patients by means of the MMPI. *Quarterly Journal of Studies on Alcohol, 26,* 238–246.

Marks, P. A., Seeman, W., & Haller, D. (1974). *The actuarial use of the MMPI with adolescents and adults*. Baltimore: Williams and Wilkins.

Matarazzo, J. D. (1986a). Computerized psychological test interpretations: Unvalidated plus all mean and no sigma. *American Psychologist, 41,* 14–24.

Matarazzo, J. D. (1986b). Response to Fowler and Butcher on Matarazzo. *American Psychologist, 41,* 96.

McKinley, J. C., Hathaway, S. R., & Meehl, P. E. (1948). The MMPI: VI. The K scale. *Journal of Consulting Psychology, 12,* 20–31.

McNair, D. M. (1978). Review of the Clinical Analysis Questionnaire (CAQ). In O. K. Buros (Ed.), *Eighth mental measurements yearbook*. Highland park, NJ: Gryphon.

Meehl, P. E. (1945). The dynamics of "structured" personality tests. *Journal of Clinical Psychology, 1,* 296–303.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and review of the evidence*. Minneapolis: University of Minnesota Press.

Meehl, P. E., & Dahlstrom, W. G. (1960). Objective configural rules for discriminating psychotic from neurotic MMPI profiles. *Journal of Consulting and Clinical Psychology, 24,* 375–387.

Megargee, E. I. (1972). *The California Psychological Inventory handbook*. San Francisco: Jossey-Bass.

Megargee, E. I., & Bohn, M. J. (1977). A new classification for criminal offenders, IV: Empirically determined characteristics of the ten types. *Criminal Justice and Behavior, 4,* 149–210.

Mesthene, E. G. (1970). *Technological change: Its impact on man and society*. New York: New American Library.

Miller, H. R., & Streiner, D. L. (1986). Differences in MMPI profiles with the norms of Colligan et al. *Journal of Consulting and Clinical Psychology, 54,* 843–845.

Millon, T. (1982). *Millon Clinical Multiaxial Inventory* (3rd ed.). Minneapolis, MN: National Computer Systems.

Millon, T. (1985a). The MCMI provides a good assessment of DSM-III disorders: The MCMI-II will prove even better. *Journal of Personality Assessment, 49,* 379–391.

Millon, T. (1985b). Response to Greer's review of the MCMI. *Journal of Counseling and Development, 63,* 631–632.

Millon, T. (1986). The MCMI and DSM-III: Further commentaries. *Journal of Personality Assessment, 50,* 205–207.

Millon, T. (1987). *Manual for the MCMI-II* (2nd ed.). Minneapolis, MN: National Computer Systems.

Mitchell, J. V. (Ed.) (1985). *The ninth mental measurements yearbook*. Lincoln: University of Nebraska Press.

Moreland, K. L. (1985). Validation of computer-based test interpretations: Problems and prospects. *Journal of Consulting and Clinical Psychology, 53,* 816–825.

Moreland, K. L. (1987). Computerized psychological assessment: What's available. In J. N. Butcher (Ed.), *Computerized psychological assessment: A practitioner's guide* (pp. 26–29).

Murray, H. A. (1938). *Explorations in personality*. Cambridge, MA: Harvard University Press.

Nelson, L. D., & Marks, P. A. (1985). Empirical correlates of infrequently occurring MMPI code types. *Journal of Clinical Psychology, 41,* 477–482.

Nichols, D. S. (1985). Review of the Minnesota report: Personnel Selection System for the MMPI. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 1008–1009). Lincoln: University of Nebraska Press.

Osborne, D. & Colligan, R. C. (1986). Linear equations for the development of non-normalized T-score tables based on the contemporary normative study of the MMPI. *Journal of Clinical Psychology, 42,* 482–484.

Patience, W. M. (1977). Description of components in tailored testing. *Behavior Research Methods and Instrumentation, 9,* 153–157.

Piotrowski, Z. A. (1980). CPR: The psychological X-ray in mental disorders. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 85–108). Norwood, NJ: Ablex.

Reise, S. P., & Waller, N. G. (in press). *Fitting the two-parameter model to personality data. Applied Psychological Measurement*.

Richards, J. S., Fine, P. R., Wilson, T. L., & Rogers, J. T. (1983). A voice-operated method for administering the MMPI. *Journal of Personality Assessment, 47,* 167–170.

Rome, H. P., Swenson, W. M., Mataya, P., McCarthy, C. E., Pearson, J. S., Keating, F. R., & Hathaway, S. R. (1962). Symposium on automation technics in personality assessment. *Proceedings of the Staff Meetings of the Mayo Clinic, 37,* 61–82.

Sanders, R. L. (1985). Computer-administered individualized psychological testing: A feasibility study. *International Journal of Man-Machine Studies, 23,* 197–213.

Sapinkopf, R. C. (1978). A computer adaptive testing approach to the measurement of personality variables. *Dissertation Abstracts International, 38* (10–B), 4993.

Schuerger, J. M., Foerstner, S. B., Serkownek, K., & Ritz, G. (1987). History and validities of the Serkownek subscales for MMPI scales 5 and 0. *Psychological Reports, 61,* 227–235.

Scissons, E. H. (1976). Computer administration of the California Psychological Inventory. *Measurement and Evaluation in Guidance, 9,* 22–25.

Sheldon, W. H. (1940). *The varieties of human physique: An introduction to constitutional psychology*. New York: Harper & Row.

Slutske, W. S., Ben-Porath, Y. S., & Butcher, J. N. (1988, March). *A real-data simulation study of adaptive MMPI administration*. Paper presented at the 23rd Annual Symposium on Recent Developments in the Use of the MMPI, St. Petersburg, Florida.

Smail, D. J. (1971). Statistical prediction and "cookbooks:" A technological confidence trick. *British Journal of Medical Psychology, 44,* 173–178.

Song, W. Z. (1981). *Application of the Minnesota Multiphasic Personality Inventory in some areas of the People's Republic of China*. Paper given at the Seventh International Conference on Personality Assessment, Honolulu, Hawaii.

Stein, K. B. (1968). The TSC scales: The outcome of a cluster analysis of the 550 MMPI items. In P. McReynolds (Ed.), *Advances in psychological assessment* (Vol. 1). Palo Alto, California: Science and Behavior Books.

Stillman, R., Roth, W. T., Colby, K. M., & Rosenbaum, C. P. (1969). An on-line computer system

for initial psychiatric inventory. *American Journal of Psychiatry, 125,* 8–11.

Stout, R. L. (1981). New approaches to the design of computerized interviewing and testing systems. *Behavior Research Methods and Instrumentation, 13,* 436–442.

Streiner, D. L., & Miller, H. R. (1986). Can a good short form of the MMPI ever be developed? *Journal of Clinical Psychology, 42,* 109–113.

Sundberg, N. D. (1985a). Review of Behaviordyne Psychodiagnostic Laboratory Service for the Minnesota Multiphasic Personality Inventory (a computer-based test interpretation). In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 1003–1005). Lincoln: University of Nebraska Press.

Sundberg, N. D. (1985b). Review of the Western Psychological Services computer based test report on the Minnesota Multiphasic Personality Inventory. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 1010–1011). Lincoln: University of Nebraska Press.

Tellegen, A. (1982). *Brief manual for the Differential Personality Questionnaire*. Minneapolis: Author.

Terman, L. M., & Miles, C. C. (1936). *Sex and personality: Studies in masculinity and femininity*. New York: McGraw-Hill.

Vale, C. D., & Keller, L. S. (1987). Developing expert computer systems to interpret psychological tests. In J. N. Butcher (Ed.), *Computerized psychological assessment: A practitioner's guide* (pp. 64–83).

Vale, C. D., Keller, L. S., & Bentz, V. J. (1986). Development and validation of a computerized interpretation system for personnel tests. *Personnel Psychology, 39,* 525–542.

Vale, C. D., & Weiss, D. J. (1975, October). *A study of computer-administered stradaptive ability testing*. (Research Report 75–5). Minneapolis: University of Minnesota, Psychometric Methods Program.

Valsiner, J. (1982). *Use of the MMPI in the Soviet Union*. Unpublished mimeographed materials.

Waller, N. B., & Reise, S. P. (in press). The development and implementation of computerized adaptive testing within the realm of normal range personality assessment. *Journal of Personality and Social Psychology*.

Weiner, H. (1956). Subtle and obvious keys for the MMPI. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine* (pp. 195–204). Minneapolis: University of Minnesota Press.

Weiss, D. J. (1973, September). *The stratified adaptive computerized ability test* (Research Report 73–3). Minneapolis: University of Minnesota, Psychometric Methods Program.

Weiss, D. J. (1985). Adaptive testing by computer. *Journal of Consulting and Clinical Psychology, 53,* 774–789.

Weiss, D. J., & Vale, C. D. (1987). Computerized adaptive testing for measuring abilities and other psychological variables. In J. N. Butcher (Ed.), *Computerized psychological assessment: A practitioner's guide* (pp. 325–343). New York: Basic Books.

Welsh, G. S. (1956). Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: University of Minnesota Press.

Widiger, T. A. (1985). Review of Millon Clinical Multiaxial Inventory. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 986–988). Lincoln: The University of Nebraska Press.

Widiger, T. A., & Sanderson, C. (1987). The convergent and discriminant validity of the MCMI as a measure of the DSM-III personality disorders. *Journal of Personality Assessment, 51,* 228–242.

Widiger, T. A., Williams, J. B. W., Spitzer, R. L., & Frances, A. (1985). The MCMI as a measure of DSM-III. *Journal of Personality Assessment, 49,* 366–378.

Wiggins, J. S. (1969). Content dimensions in the MMPI. In J. N. Butcher (Ed.), *MMPI: Research developments and clinical applications*. New York: McGraw-Hill.

Williams, T. A. (1977). Computer technology in mental health care: "Toys" or tools? *Behavior Research Methods and Instrumentation, 9,* 108–109.

Woodworth, R. S. (1920). *Personal Data Sheet*. Chicago: Stoelting.

Zuckerman, M. (1985). Review of Sixteen Personality Factor Questionnaire. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 1392–1394). Lincoln: University of Nebraska Press.

4.14
Objective Personality Assessment with Adults

JAMES N. BUTCHER and JEANETTE TAYLOR
University of Minnesota, Minneapolis, MN, USA
and
G. CYNTHIA FEKKEN
Queen's University, Kingston, ON, Canada

4.14.1 INTRODUCTION 404
4.14.1.1 Intelligence and Insight 404
4.14.1.2 Veridicality of Self-report 405
4.14.1.3 Personality Stability 405
4.14.1.3.1 Definitions 405
4.14.1.3.2 Influences on personality stability 406
4.14.2 ASSESSING ADULTS IN CLINICAL SETTINGS 407
4.14.2.1 The Minnesota Multiphasic Personality Inventory-Revised 407
4.14.2.1.1 Origin of the MMPI/MMPI-2 408
4.14.2.1.2 Measurement dimensions 408
4.14.2.1.3 Recent validity research for the MMPI-2 409
4.14.2.2 Basic Personality Inventory 413
4.14.2.3 Personality Assessment Inventory 415
4.14.3 SPECIALIZED OR FOCUSED CLINICAL ASSESSMENT MEASURES 416
4.14.3.1 Millon Clinical Multiaxial Inventory 416
4.14.3.2 The Beck Depression Inventory 418
4.14.3.3 The State-Trait Anxiety Inventory 418
4.14.3.4 Whitaker Index of Schizophrenic Thinking 419
4.14.4 NORMAL RANGE PERSONALITY ASSESSMENT 420
4.14.4.1 Objective Personality Measures in Research 420
4.14.4.1.1 The Five Factor Model (the Big Five) 420
4.14.4.1.2 NEO Personality Inventory 420
4.14.4.1.3 Multidimensional Personality Questionnaire 420
4.14.4.1.4 Personality Research Form 421
4.14.4.1.5 Sixteen Personality Factor Test 421
4.14.4.1.6 California Psychological Inventory 421
4.14.4.2 Objective Personality Measures in Educational/Vocational Assessment 422
4.14.4.2.1 FFM 422
4.14.4.2.2 NEO-PI 422
4.14.4.2.3 CPI 422
4.14.4.2.4 MMPI/MMPI-2 422
4.14.4.3 Personnel Screening 423
4.14.4.4 Other Personality Measures in Personnel Selection 423
4.14.4.5 The 16PF 423
4.14.4.6 FFM 423

4.14.4.7 NEO-PI 424
4.14.4.7.1 CPI 424
4.14.5 SUMMARY 424
4.14.6 REFERENCES 425

4.14.1 INTRODUCTION

Objective personality tests form a standard part of most applied psychologists' toolboxes when it comes to measuring personality. In contrast to projective techniques and to subjective approaches to personality assessment, objective personality questionnaires are made up of relatively unambiguous stimuli or items; offer the respondent relatively restricted response options; and present a scoring scheme that involves few, if any, scoring judgments, resulting in high scoring reliability (Wiggins, 1973). Most objective personality assessment instruments used by clinicians and researchers employ a self-report format. In contemporary psychology, the Minnesota Multiphasic Personality Inventory (MMPI) and the MMPI-2 are the most widely used personality assessment instruments in both clinical and research settings (Butcher and Rouse, 1996). The next most popular instruments used to assess personality, the Rorschach and the Thematic Apperception Test, are projective (not objective), but each still requires the clinician to obtain self-reported behavior. In essence, the clinician is dependent on the client's ability and willingness to make accurate self-reports when assessing personality with standardized instruments. It is, therefore, important to determine whether adults being assessed can competently reveal information about their personalities through self-report instruments. In order to address this question, the client's intelligence and insight in sharing self-information and the success of personality tests for eliciting self-information that can be externally verified need to be considered. We will examine each of these considerations below.

4.14.1.1 Intelligence and Insight

The existing literature on intelligence and personality is largely focused on the predictive relationship between the two constructs. However, clinicians are generally not concerned about the ability of a personality profile to predict an intelligence quotient or vice versa. Instead, most are concerned with the more basic question of how intellectual functioning impacts the ability of a client to self-disclose on an objective personality test. No literature exists which directly addresses the issue of the relationship between a client's level of intelligence and his or her ability to competently reveal information about personality through a self-report instrument.

Objective personality measures leave the issue of the client's intellectual functioning in the hands of the clinician. None of the widely used personality assessment instruments require a pre-screening for normal or above normal intelligence. However, most self-report personality instruments require that the examinee have a minimum reading level competency (often expressed as a minimum grade level). Thus, a minimum level of school achievement is implicit in the validity of a client's score, and it follows that the clinician should use care when administering a self-report instrument to a client with known or suspected intellectual impairment. It is also essential for the clinician to be aware of the educational range of the standardization sample for any self-report measure, and to use caution when interpreting protocols from clients who fall outside of that range.

The clinician needs to be aware of a client's intellectual functioning as it may influence the client's ability to provide information about his or her personality on a self-report measure. If a client appears to be low in intellectual functioning, then the results of self-report measures should be interpreted with caution and perhaps supplemented with reports from other sources.

While the capacity for insight cannot readily be extrapolated from an intelligence quotient in a linear fashion, insight into one's problems might be discernible from the personality profile itself. As Butcher and Rouse (1996) note in their recent review of personality research, several of the widely used objective personality measures have attitudinal measures that aid the clinician in evaluating the client's disclosure capability. Crookes and Buckley (1976) found a relationship between the Eysenck Personality Inventory Lie (L) scale score and diagnosis among both psychiatric inpatients and outpatients. They found that high L scale scores were positively related to diagnosis of disorders associated with low insight. They concluded that the L score (a validity indicator found on several objective personality instruments) is highly related to a person's awareness of his or her behavior. Moreover, this finding is in agreement with the interpretive guidelines for the MMPI-2 L scale, which suggests the general utility of this measure at appraising a client's self-awareness.

The quality of insight that a client brings to the assessment situation may greatly influence the validity of the assessment. Some instruments are designed to assess clients through behavioral items that require low insight or self-observation. These items are specifically selected because of their ability to predict relevant criteria and may appear at face value totally unrelated to the particular aspects of personality functioning intended to be predicted. This highlights one advantage of using objective personality measures in clinical settings: clients who may yield poor information during a clinical interview that requires self-evaluation and/or introspection may be fully capable of providing valuable clinical information when a personality instrument comprising empirically selected items is employed.

4.14.1.2 Veridicality of Self-report

Aside from the issues of intellectual functioning and insight the utility of self-report measures can be assessed by examining their success at corroborating independent facts about a client. For example, in an early study of the MMPI, Payne and Wiggins (1972) examined the relationship between content-based profiles and external descriptors of a large group of psychiatric inpatients. Interpretation of content-based profiles is based on combinations of "obvious" test items which correspond to traditional self-report instruments (e.g., "I am a high-strung person"). That is, the client endorses the item as a means of directly relaying information about him- or herself. The authors found that the patients' self-reported MMPI profiles matched quite well with interview report and external observation.

In a similar vein, Koss and Butcher (1973) found that psychiatric inpatients identified (through their observed behavior and presenting symptoms) as belonging to one of six operationally defined crisis situations could be discriminated from one another on the basis of their endorsed MMPI content. The authors interpreted their results as providing evidence for the competence and willingness of adults in clinical settings to reflect information accurately about their personality and psychological functioning through self-report test items. Similar evidence of a positive relationship between self-reported personality measures and external criteria has been found for certain scales from the California Psychological Inventory (CPI; e.g., Hindelang, 1972), the Millon Clinical Multiaxial Inventory (MCMI; e.g., Jaffe and Archer, 1987), and the Basic Personality Inventory (BPI; Holden, Fekken, Reddon, Helmes, & Jackson, 1988), to name three.

Although the research literature shows that people can competently disclose information about their personalities through standardized, objective, self-report measures, some individuals, however, may not be motivated to cooperate in their psychological assessment. That is, some respondents, because of a need or motivation to present themselves in a particular way, do not respond in a truthful, open manner. The most common examples of individuals being motivated to appear different from the way they actually are in today's assessment settings are applicants for employment, parents involved in custody disputes, and other individuals being evaluated as part of a court case. A self-report assessment instrument must contain an effective means of identifying respondents who are dissimulating. Perhaps the most common way of detecting whether people are not motivated to report accurately about themselves is via validity scales or control scales. In some assessment situations the validity scales provide the most important and useful information about the client. A self-report personality measure without effective validity scales is too limited to operate across a broad range of assessment situations. Furthermore, tests that possess validity indicators, such as the L scale on the MMPI/MMPI-2, offer the clinician information regarding both the client's willingness to cooperate with the assessment and his or her level of insight. These points argue in favor of the continued use of standardized, objective self-report measures in the assessment of personality in clinical settings.

4.14.1.3 Personality Stability

One of the assumptions of objective personality assessment is that the personality characteristic being measured is stable. Various major studies have come to the conclusion that personality changes little in adulthood, particularly after age 30 (Conley, 1985; Finn, 1986; McCrae & Costa, 1990; Schuerger, Zarrella, & Hotz, 1989). Typical levels of stability on personality tests are in the 0.5–0.7 range. This generalization bears some comment.

4.14.1.3.1 Definitions

What exactly is meant by stable responding? On an objective personality questionnaire, a set of responses that are consistent over time may be defined in at least three ways: (i) as an identical scale score, (ii) as a set of identical item

responses, or (iii) as a scale score that signifies the same relative standing on the personality characteristic.

Conclusions about the stability of personality are most commonly based on scale scores. That is, the stability of personality estimates result from comparing scale scores obtained at two points in time. Presumably the same scale score could result even if the test respondent endorsed somewhat different subsets of items. The consistency of responses to individual test items has been studied as a meaningful individual differences variable and as an index for establishing the interpretability of individual response protocols. Item response consistency has not been examined as a definition of longitudinal personality stability *per se*.

Many researchers have focused on changes in average scale scores. Their goal is to understand normative changes in personality. On the other hand, some researchers examine personality stability by correlating scale scores. Thus, when coefficients of personality stability are reported to be in the 0.5–0.7 range, this should be understood to be a statement about relative, and not absolute, scale score stability.

4.14.1.3.2 Influences on personality stability

Although the typical stability of personality scale scores is quite high, there are nonetheless particular variables that affect score stability in predictable ways. The first consideration is the influence of instrument characteristics. Specifically, the average number of items per scale and the homogeneity of the scales are the important predictors of scale score stability (Finn, 1986; Schuerger et al., 1989). Variance associated with specific instruments (e.g., the MMPI vs. the Sixteen Personality Factor Test [16PF]) apparently does not contribute to greater personality stability beyond the effects of scale length and homogeneity (Schuerger et al., 1989). Such findings underscore the need for test users to select reliable measures, particularly if test data are intended to evaluate long-term personality change.

Second, estimates of personality stability are affected by the length of the retest interval. Not surprisingly, personality appears to more stable over short intervals than long ones. In particular, score stability drops over the first year before it levels off (Schuerger et al., 1989; Windle, 1954) and then stays stable at the same level for very long periods of time (McCrae & Costa, 1990).

One simple reason for relatively high short-term personality stability may be that people remember and repeat their responses to the items on a questionnaire. Lumsden (1977), however, proposed a substantive rather than a methodological explanation. He argued that particular personality characteristics "swell" into prominence over the short term and then fade again. For a student entering college, for example, "independence" may become a salient dimension and hence, self-reports on this dimension are likely to have exaggerated consistency over the (short) time period when "independence" is prominent.

A third influence on personality stability has to do with the operationalization of personality. Most objective personality questionnaires are intended to measure traits that are by definition stable and enduring. The construction of such questionnaires favors selection of items that will yield stable scores. In addition, the overall instructions as well as the wording of specific items prompt people to make broad generalizations about themselves and to downplay variations. Thus, personality may appear to be stable because the constructs of personality are explicitly conceptualized and measured as stable. There are certainly empirical studies that challenge the idea of personality stability by providing evidence of developmental change over long time periods based on personality constructs and measures that expressly incorporate the notion of change (e.g., Whitbourne, Zuschlag, Elliot, & Waterman, 1992).

Fourth, some personality constructs are associated with more stability than others. Schuerger et al.'s (1989) data show that measures of psychopathology have less scale score stability than measures of "normal" personality characteristics. Perhaps "normal" personality constructs are more crystallized than psychopathological constructs and thus less error is built into the construct itself. Similarly, psychopathological constructs may be more confounded with the effects of response style variance such as social desirability and acquiescence than normal personality constructs, again compounding the noisiness of the construct.

Many researchers have tried to highlight differences in stability among the specific constructs that they chose to study. For example, Finn (1986) demonstrated that mood-related constructs (e.g., depression) had comparatively low levels of consistency, whereas Helson and Moane (1987) found that constructs related to socialization had high stability. There are shortcomings to such an approach. The identification of those dimensions on which people are changeable (i.e., states) as distinct from the more permanent individual differences dimensions needs to be tackled systematically if a comprehensive list is to be obtained (Cattell, 1963).

The interpretation of differences in stability between constructs is further complicated by age and cohort effects. Studies such as that of Helson and Moane (1987) estimate personality stability using information obtained from a single group of subjects who have been assessed at different points in time. There are limits to the generalizability of this information to other groups of people. Other studies such as that of Finn (1986) use basically a cross-sectional approach. They assess two or more groups of subjects having different ages at a single time. Here it is difficult to attribute differences in the stability of particular personality constructs to age as opposed to generational factors. More complicated designs may be needed in order to understand such age and cohort effects (Conley, 1985). At the very least, any generalizations about which personality constructs are more or less stable need to be carefully qualified.

A fifth influence on personality stability has to do with features of the persons being evaluated. Test respondents exhibit remarkable individual differences in personality stability (Assendorp, 1992; McCrae & Costa, 1990). Do some subgroups of people show more personality stability? Across studies, patient and prisoner samples show less personality test score stability than "normals" (Schuerger et al., 1989). Perhaps "non-normals" have systematically higher scale scores than "normals" and, hence, their retest scores show regression toward the mean (Windle, 1954). Alternatively, "non-normal" groups also may be more likely than normal groups to seek actively to change personality.

The other two person variables that have received attention in the literature on personality stability are age and sex differences. Some researchers report that older adults are more stable than younger adults on many traits (Finn, 1986; Schuerge et al., 1989) although others argue that stability differences after about age 30 are minimal (McCrae & Costa, 1990). Finn (1986) has asked whether older adults might just be more rigid in their self-perceptions. However, the picture of personality stability that results from self-reports does tend to be substantiated by ratings on personality information collected from others, such as spouses (McCrae & Costa, 1990). Thus, personality may indeed be more stable in older adults.

Many studies in the area of personality stability sample exclusively men or women. Yet there seems to be little evidence of an overall difference in the level of personality stability for men and women (Schuerger et al., 1989). Nonetheless, researchers interested in normative change often argue that men (Finn, 1986) and women (Helson & Moane, 1987) may be expected to change on different constructs as a function of distinct life paths or societal pressures. Few studies have been explicitly designed to compare sex differences in personality stability.

Overall, personality shows considerable stability in adulthood. The objective measures traditionally used in personality assessment may predispose us to find evidence of stability as a function of the kinds of constructs, instructions, and test construction techniques associated with these measures. Aspects of the test instrument itself, such as length or homogeneity, may moderate stability estimates, as could the length of the interval between personality assessments. Samples of "normals" and of older adults may exhibit more personality stability; sex differences in general personality stability appear to be minimal. The assumption of personality stability, at least over the short run, is central to objective personality assessment if test scores are going to predict relevant behaviors and have an impact on planning how to manage that behavior.

4.14.2 ASSESSING ADULTS IN CLINICAL SETTINGS

In this section we will provide an overview of a number of objective personality inventories that have been designed to assess adults in clinical settings. More detail on the use of the MMPI-2 will be given because it is the most widely researched and used measure. Two other measures that have recently been published to measure essentially the same clinical problem areas as the MMPI-2, BPI (Jackson, 1989) and the Personality Assessment Inventory (PAI) (Morey, 1991) will be briefly described. In the next section, several specialized measures will then be surveyed that have been developed to measure more specific clinical problems or behaviors: the MCMI to assess personality disorders; the Beck Depression Inventory (BDI; Beck, Steer, & Garbin, 1988) to assess anxiety; the State–Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, & Lushene, 1970); and the Whitaker Index of Schizophrenic Thinking (WIST; Whitaker, 1973).

4.14.2.1 The Minnesota Multiphasic Personality Inventory-Revised

The MMPI-2 is the most widely used personality test with adults (Lubin, Larsen, & Matarazzo, 1984; Watkins, 1996). Keilen and Bloom (1986) found that the MMPI/MMPI-2 was the most frequently used test in custody evaluations. Developed originally in the late

1930s by Starke Hathaway and J. C. McKinley and redeveloped by Butcher, Dahlstrom, Graham, Tellegen, and Kaemmer (1989), this instrument provides a comprehensive survey of personality characteristics and clinical problems. In the original MMPI, a strictly empirical scale construction approach was followed to develop scales that would assess a patient's probable "membership" in a clinical group. An extensive amount of research has been published on the effectiveness of these measures at predicting and describing problems in adults. Research on the original version of the MMPI covered a very broad range of people—psychiatric and medical patients, substance abusers, incarcerated felons, and many other clinical groups.

In addition, a very broad range of "normals" have been studied including applicants for various jobs such as airline pilots, US Navy submariner crew members, and police and security personnel. Moreover, the instrument came to be widely employed as a personality research instrument. Butcher and Rouse (1996), in a survey of 20 years of research in clinical assessment, found that the MMPI/MMPI-2 is the most widely researched instrument with nearly twice the number of articles as the seven next leading tests. What has made the MMPI/MMPI-2 the most widely used assessment technique in the personality area? We will examine the make-up, utility, and limitations of the instrument in assessing adults.

4.14.2.1.1 Origin of the MMPI/MMPI-2

A growing dissatisfaction with subjective methods of evaluating patients in clinical situations (such as interviews and projective tests) led the original MMPI developers, Hathaway (a psychologist) and McKinley (a psychiatrist), to experiment with an objective method of clinical problem assessment. They accumulated a large number of items (symptoms, beliefs, attitudes, etc.) from tests and case material and administered them to a large group of "normal" people to serve as a comparison group. They then administered the items to well-defined and homogeneous groups of clinical patients. They empirically contrasted the responses of the patient groups with the normals to obtain items that significantly separated the groups. These items were then combined into scales for the various clinical problem areas such as depression, schizophrenia, and so forth. Once derived, the clinical scales were normed on the non-patient sample and *T* scores were developed to enable the test interpreter to determine how extreme a particular person's score was on a given scale. All of the scales were normed so that they had a mean score of 50 and a standard deviation of 10. These standard score distributions then allowed the scores to be plotted on a profile so that the interpreter would have a visual picture of how extreme a particular score was when compared with the normal. Their empirical approach produced highly valid and effective scales that predicted or described the likelihood that a person's score on a scale was in the clinical range or similar to the patient groups.

The original MMPI clinical scales have undergone very substantial study and cross-validation since their publication. They have become a standard means of objective symptom classification since their development. The clinical scales and configurations of scales, referred to as code types, have undergone substantial documentation as an objective classification schema—often referred to as "cookbooks." Researchers, for example, have cataloged the behavioral characteristics associated with their test indices allowing for automatic interpretation. That is, when a particular scale score or cluster of scales are obtained, then a well-validated set of behavioral descriptions—known as descriptors—are applied. Their objective classification approach to interpretation fostered the development of computer-based interpretation methods so popular today (Butcher, 1995; Butcher et al., 1998).

4.14.2.1.2 Measurement dimensions

(i) Validity scales

As noted earlier, it is essential to any self-reported personality assessment to appraise carefully possible invalidating conditions or circumstances. Structural elements of the test administration require evaluation to determine if extratest factors influenced the item responses. For example, were the instructions clearly presented, was the individual able to read and comprehend the items?

The MMPI-2 contains a number of scales and indexes that provide the test interpreter with information on the individual's cooperativeness and honesty in responding to the items (see Table 1 for a summary of the validity or control scales for the MMPI-2). A clear picture is obtained as to whether the person has attempted to present a false picture on the test. These scales are of several types. First, noncontent-oriented measures can provide information on whether the person was inconsistent in responding (the True Response Inconsistency or TRIN and Variable Response Inconsistency or VRIN scales). Additionally, the Cannot Say or "?" scale provides information about the person's cooperation in completing the items. Two other indexes that can provide clues to uncooperative test-taking behavior are the percentage of true and percentage of false endorsement. Records with a nearly all true or all false response pattern suggest uncooperativeness in completing the task.

Three scales have been developed to provide information about the tendency on the part of some people to claim excessive virtue or to deny problems. The L scale assesses an unsophisticated tendency to claim excessive virtue; the K scale measures test defensiveness; and the S scale (Superlative Self-presentation) assesses the tendency of some people to present themselves in a highly favorable manner (Butcher & Han, 1995).

Three scales have been developed to assess the tendency that some respondents have to claim excessive mental health symptoms. The original F scale developed by Hathaway and McKinley (1940) assesses excessive symptom checking by attending to extreme or rare responses. Those who endorse a large number of rare items are thought to be presenting an unselective or exaggerated complaint pattern. High scores on F have been associated with malingering (Berry, Baer, & Harris, 1991; Schretlen, 1988).

The F(b) scale, an infrequency scale that covers extreme items in the back of the MMPI-2 items operates much like the original F scale (Butcher et al., 1989). The newest validity measure developed by Arbisi and Ben-Porath (1995) is an extremely valuable type of infrequency scale that differs from the original F scale in an important way: the F(p) scale assesses extreme responding in a clinical sample. The F(p) scale assesses infrequent responding in a clinical setting and thereby, when extreme, suggests malingering of psychiatric symptoms.

(ii) Clinical scales

The traditional MMPI clinical scales (Hathaway & McKinley, 1940) have been updated in the MMPI-2 (Butcher et al., 1989) and have been extensively revalidated in their revised form (Archer, Griffin, & Aiduk, 1995; Butcher & Williams, 1992; Graham & Ben-Porath, 1990; Graham & Butcher, 1988). The MMPI-2 clinical scales are empirically derived measures of several well-established clinical patterns such as hypochondriasis, depression, paranoia, and schizophrenia (see Table 2).

These scales have been well studied over more than 50 years. They were kept nearly intact in MMPI-2 because of their descriptive power and ability to generalize across groups.

Two scales on the clinical profile, Mf and Si, are not empirically derived scales as were the original clinical scales but were included on the instrument to assess personality characteristics of masculine–feminine interests and social introversion.

(iii) Content scales

The expanded item pool of the MMPI-2 allowed for the development of a new set of content-based scales to assess an expanded array of problems compared with the traditional clinical measures. The 15 content scales (see Table 3) were developed according to a multimethod/multistage strategy. The 567 items were subjected to a content analysis in order to derive homogeneous content groups rationally. This approach is not as theoretically blind as it seems, since the original MMPI content had been well delineated by Wiggins (1969) and many of these constructs were still available in a modified or reduced form in the revised instruments. Moreover, the MMPI-2 committee also wrote new items to cover contents that were limited or not available in MMPI-2, for example, suicide ideation, substance abuse, and type A behavior. It therefore became possible to develop scales for a number of problem areas that have clinical relevance and are sufficiently large to assess reliably. Following provisional derivation, other scale construction strategies such as internal consistency, item scale correlation, and external validation were followed to define and improve the item groups.

Content scales provide a different type of information to the clinical interpreter's resources. These scales are viewed as direct communications between the client and the clinician. The obvious content on the scales enables the client to address problems he or she has that are considered important to address in the clinical intervention.

In addition to their value as summaries or themes considered pertinent by the patient the content scales have clearly established external validity. A number of studies have provided data on the external validity of the content scales (Ben-Porath, Butcher, & Graham, 1991; Butcher, Graham, Williams, & Ben-Porath, 1990).

4.14.2.1.3 Recent validity research for the MMPI-2

An empirically derived and based instrument requires substantial research validation in order to be useful. The MMPI-2 has been substantially researched and validated. The fact that the MMPI-2 clinical scales are the same items as in the original instrument means that all of the research on the traditional measures applies with the revised form.

Table 1 Personality characteristics associated with validity indicator elevations.

? Cannot say score

The total number of unanswered items. A defensive protocol with possible attenuation of scale scores is suggested if the ? raw score is more than 35.

L (Lie) scale

A measure of an unsophisticated or self-consciously "virtuous" test-taking attitude. Elevated scores (above 70 *T*) suggest that the individual is presenting himself or herself in an overly positive light, attempting to create an unrealistically favorable view of his or her adjustment.

F (Infrequency) scale

A high score (*T* above 90) suggests an exaggerated pattern of symptom checking that is inconsistent with accurate self-appraisal and suggests confusion, disorganization, or actual faking of mental illness. *T* scores above 100 invalidate the profile.

F(B) scale

A second infrequency scale that appears toward the back of the MMPI-2 item pool is used to assess exaggerated responding at the end of the test. This scale operates much like the original MMPI F scale in detecting malingering, random responding, or motivation to exaggerate symptoms.

F(p) scale

The Psychopathology Infrequency Scale F(p) was developed by Arbisi and Ben-Porath (1995) to assess infrequent responding in psychiatric settings. This scale is valuable in appraising the tendency for some people to exaggerate mental health symptoms in the context of patients with genuine psychological disorder.

K (Defensiveness) scale

Measures an individual's willingness to disclose personal information and discuss his or her problems. High K scores (*T* above 65) reflect an uncooperative attitude and an unwillingness or reluctance to disclose personal information. Low scores (*T* below 45) suggest openness and frankness. This scale is positively correlated with intelligence and educational level, which should be taken into account when interpreting the scores.

Variable response inconsistency

The VRIN scale consists of 49 pairs of specially selected items. The members of each VRIN item pair have either similar or opposite content; each pair is scored for the occurrence of an inconsistency in responses to the two items. The scale score is the total number of item pairs answered inconsistently. High VRIN scores are a warning that a test subject may have been answering the items in the inventory in an indiscriminate manner, and raise the possibility that the protocol may be invalid and that the profile is essentially uninterpretable.

True response inconsistency

The TRIN is made up of 20 pairs of items that are opposite in content. If a subject responds inconsistently by answering True to both items of certain pairs, one point is added to the TRIN score; if the subject responds inconsistently by answering False to certain item pairs, one point is subtracted. A very high TRIN score indicates a tendency to give True answers to the items indiscrimately ("acquiescence"), and a very low TRIN score indicates a tendency to answer False indiscriminately ("nonacquiescence"). (Negative TRIN scores are avoided by adding a constant to the raw score.) Very low or very high TRIN scores are a warning that the test subject may have been answering the inventory indiscriminately so that the profile may be invalid and uninterpretable.

S scale

A new MMPI-2 scale was developed in an effort to improve the assessment of highly virtuous responding. The subjects used in the development of the scale were 274 male airline applicants and the MMPI-2 normative sample ($N = 1138$ men and 1462 women). The S scale was initially developed by examining item response differences between airline pilot applicants, who tend to engage in superlative self-description in order to impress examiners, and normative men from the MMPI-2 restandardization sample. The scale was refined by using internal consistency methods to ensure high-scale homogeneity. A factor analysis of the resulting 50 item scale (S) yielded five factors named: Beliefs in "Human Goodness"; Serenity; Contentment with Life; Patience and Denial of Irritability and Anger; and Denial of Moral Flaws. Linear *T* score conversion tables were computed for both men and women separately and combined (unisex) using the MMPI-2 restandardization data sets. The S scale was shown to have a number of behavioral correlates reflecting the presentation of oneself as a well-controlled, problem-free person.

Table 2 Empirical correlates for MMPI-2 clinical scales.

Scale 1 (Hypochondriasis)

High-scoring people show: excessive bodily concern; somatic symptoms that tend to be vague and undefined; fatigue, pain, weakness; manifest anxiety; selfish, self-centered, and narcissistic behavior; pessimistic, defeatist, cynical outlook on life; dissatisfied and unhappy; make others miserable with whining, complaining behavior; demanding and critical of others; expresses hostility indirectly; rarely act out; dull, unenthusiastic, unambitious; ineffective in oral expression, longstanding health concerns; function at a reduced level of efficiency without major incapacity; not very responsive to therapy, tend to terminate therapy when therapist is seen as not giving enough attention and support; tend to seek medical solutions to life problems.

Scale 2 (Depression)

High-scoring people show: depressed, unhappy, and dysphoric behavior; they are pessimistic and self-deprecating; tend to feel guilty; report being sluggish; have somatic complaints such as weakness, fatigue, and loss of energy; are agitated, tense, highly-strung, and irritable; prone to worry; lack self-confidence; feel useless and unable to function; feels like a failure at school or on the job; introverted, shy, retiring, timid, and seclusive; aloof, maintain psychological distance; avoids interpersonal involvement; cautious, and conventional; has difficulty making decisions; nonaggressive; overcontrolled, deny impulses; make concessions to avoid conflict; motivated for therapy.

Scale 3 (Hysteria)

High-scoring people show: poor response to stress; they avoid responsibility through development of physical symptoms; have headaches, chest pains, weakness, and tachycardia, anxiety attacks; symptoms appear and disappear suddenly; lack insight about causes of symptoms; lack insight about own motives and feelings; lack anxiety, tension, and depression; rarely report delusions, hallucinations, or suspiciousness; psychologically immature, childish, and infantile; self-centered, narcissistic, and egocentric; expect attention and affection from others; use indirect and devious means to get attention and affection; do not express hostility and resentment openly; socially involved; friendly, talkative, and enthusiastic; superficial and immature in interpersonal relationships; show interest in others for selfish reasons; occasionally act out in sexual or aggressive manner with little apparent insight; initially enthusiastic about treatment; respond well to direct advice or suggestion; slow to gain insight into causes of own behavior; resistant to psychological interpretations.

Scale 4 (Psychopathic deviate)

High-scoring people show: antisocial behavior; rebellious toward authority figures; stormy family relationships; blame parents for problems; history of under-achievement in school; poor work history; marital problems; impulsive; strive for immediate gratification of impulses; do not plan well; act without considering consequences of actions; impatient; limited frustration tolerance; poor judgment; take risks; do not profit from experience; immature, childish, narcissistic, self-centered, and selfish; ostentatious, exhibitionistic; insensitive; interested in others in terms of *how* they can be used; likeable and usually create a good first impression; shallow, superficial relationships, unable to form warm attachments; extroverted, outgoing; talkative, active, energetic, and spontaneous; intelligent; assert self-confidence; have wide range of interests; lack definite goals; hostile, aggressive; sarcastic, cynical; resentful, rebellious; act out; antagonistic; aggressive outbursts, assaultive behavior; little guilt over negative behavior; may feign guilt and remorse when in trouble; free from disabling anxiety, depression, and psychotic symptoms; likely to have personality disorder diagnosis (antisocial or passive–aggressive); prone to worry; dissatisfied; show absence of deep emotional response; feel bored and empty; poor prognosis for change in therapy; blame others for problems; intellectualize; may agree to treatment to avoid jail or some other unpleasant experience but are likely to terminate before change is effected.

Scale 5 (Masculinity–femininity)

MALES

Scores of T over 80: Show conflicts about sexual identity; insecure in masculine role; effeminate; aesthetic and artistic interests; intelligent and capable; value cognitive pursuits; ambitious, competitive, and persevering; clever, clear-thinking, organized, logical; show good judgment and common sense; curious; creative, imaginative, and individualistic in approach to problems; sociable; sensitive to others; tolerant; capable of expressing warm feelings toward others; passive, dependent, and submissive; peace-loving; make concessions to avoid confrontations; good self-control; rarely act out. (The interpretation of high 5 scores should be tempered for males with advanced academic degrees.)

High T score between 70 and 79: May be viewed as sensitive; insightful; tolerant; effeminate; showing broad cultural interests; submissive, passive. (In clinical settings, the patient might show sex role confusion; or heterosexual adjustment problems.)

Table 2 (continued)

Low T < 35: "Macho" self-image, present self as extremely masculine; overemphasize strength and physical prowess; aggressive, thrill-seeking, adventurous, and reckless; coarse, crude, and vulgar; harbor doubts about own masculinity; have limited intellectual ability; narrow range of interests; inflexible and unoriginal approach to problems; prefer action to thought; are practical and nontheoretical; easy-going, leisurely and relaxed; cheerful, jolly, humorous; contented; willing to settle down; unaware of social stimulus value; lack insight into own motives; unsophisticated.

FEMALES
Scores of T over 70: Reject traditional female roles and activities; masculine interests in work, sports, hobbies; active, vigorous, and assertive; competitive, aggressive, and dominating; coarse, rough, and tough; outgoing, uninhibited, and self-confident; easy-going, relaxed, balanced; logical, calculated; unemotional, and unfriendly.

Low T < 35: Describe self in terms of stereotyped female role; doubts about own femininity; passive, submissive, and yielding; defer to males in decision-making; self-pity; complaining, fault finding; constricted; sensitive; modest; idealistic. (This interpretation for low 5 females does *not* apply for females with postgraduate degrees.)

Scale 6 (Paranoia)

Extremely high-scoring people (T > 80) show: frankly psychotic behavior; disturbed thinking; delusions of persecution and/or grandeur; ideas of reference; feel mistreated and picked on; angry and resentful; harbor grudges; use projection as defense; most frequently diagnosed as schizophrenia or paranoid state.

Moderate elevations (= 65–79 for males; 71–79 for females): Paranoid predisposition; sensitive; overly responsive to reactions of others; feel they are getting a raw deal from life; rationalize and blame others; suspicious and guarded; hostile, resentful, and argumentative; moralistic and rigid; overemphasizes rationality; poor prognosis for therapy; do not like to talk about emotional problems; difficulty in establishing rapport with therapist.

Extremely low (T < 35): should be interpreted with caution. In a clinical setting, low 6 scores, in the context of a defensive response set, may suggest frankly psychotic disorder; delusions, suspiciousness, ideas of reference; symptoms less obvious than for high scorers; evasive, defensive, guarded; shy, secretive, withdrawn.

Scale 7 (Psychasthenia)

High-scoring people show: anxious, tense, and agitated; high discomfort; worried and apprehensive; high strung and jumpy; difficulties in concentrating; introspective, ruminative; obsessive, and compulsive; feel insecure and inferior; lack self-confidence; self-doubting, self-critical, self-conscious, and self-derogatory; rigid and moralistic; maintain high standards for self and others; overly perfectionistic and conscientious; guilty and depressed; neat, orderly, organized, and meticulous; persistent; reliable; lack ingenuity and originality in problem solving; dull and formal; vacillates; are indecisive; distort importance of problems, overreact; shy; do not interact well socially; hard to get to know; worry about popularity and acceptance; sensitive, physical complaints; shows some insight into problems; intellectualize and rationalize resistant to interpretations in therapy; express hostility toward therapist; remain in therapy longer than most patients; makes slow but steady progress in therapy.

Scale 8 (Schizophrenia)

Very high scorers (= over 80–90) show: blatantly psychotic behavior; confused, disorganized, and disoriented; unusual thoughts or attitudes; delusions; hallucinations; poor judgment.

High (65–79): schizoid lifestyle; do not feel a part of social environment; feel isolated, alienated, and misunderstood; feel unaccepted by peers; withdrawn, seclusive, secretive, and inaccessible; avoid dealing with people and new situations; shy, aloof, and uninvolved; experience generalized anxiety; resentful, hostile, and aggressive; unable to express feelings; react to stress by withdrawing into fantasy and daydreaming; difficulty separating reality and fantasy; self-doubts; feel inferior, incompetent, and dissatisfied; sexual preoccupation, and sex role confusion; nonconforming, unusual, unconventional, and eccentric; vague, long-standing physical complaints; stubborn, moody, and opinionated; immature, and impulsive; highly-strung; imaginative; abstract, vague goals; lack basic information for problem-solving; poor prognosis for therapy; reluctant to relate in meaningful way to therapist; stay in therapy longer than most patients; may eventually come to trust therapist.

Scale 9 (Hypomania)

High-scoring people (T > 80) show: overactivity; accelerated speech; may have hallucinations or delusions of grandeur; energetic and talkative; prefer action to thought; wide range of interest; do not utilize energy wisely; do not see projects through to completion; creative, enterprising, and ingenious; little interest in routine or detail; easily bored and restless; low frustration tolerance; difficulty in inhibiting expression of impulses; episodes of irritability, hostility, and aggressive outbursts; unrealistic, unqualified optimism; grandiose aspirations; exaggerates self-worth and self-importance; unable to see own limitations; outgoing, sociable, and

Table 2 (continued)

gregarious; like to be around other people; create good first impression; friendly, pleasant, and enthusiastic; poised, self-confident; superficial relationships; manipulative, deceptive, unreliable; feelings of dissatisfaction; agitated; may have periodic episodes of depression; difficulties at school or work, resistant to interpretations in therapy; attend therapy irregularly; may terminate therapy prematurely; repeat problems in stereotyped manner; not likely to become dependent on therapists; becomes hostile and aggressive toward therapist.

Moderately elevated scores (T > 65, LE 79): Over-activity; exaggerated sense of self-worth; energetic and talkative; prefer action to thought; wide range of interest; do not utilize energy wisely; do not see projects through to completion; enterprising, and ingenious; lack interest in routine matters; become bored and restless easily; low frustration tolerance; impulsive; has episodes of irritability, hostility, and aggressive outbursts; unrealistic, overly optimistic at times; shows some grandiose aspirations; unable to see own limitations; outgoing, sociable, and gregarious; like to be around other people; create good first impression; friendly, pleasant, and enthusiastic; poised, self-confident; superficial relationships; manipulative, deceptive, unreliable; feelings of dissatisfaction; agitated; view therapy as unnecessary; resistant to interpretations in therapy; attend therapy irregularly; may terminate therapy prematurely; repeat problems in stereotyped manner; not likely to become dependent on therapists; become hostile and aggressive toward therapist.

Low scorers (T below 35): Low energy level; low activity level; lethargic, listless, apathetic, and phlegmatic; difficult to motivate; report chronic fatigue, physical exhaustion; depressed, anxious, and tense; reliable, responsible, and dependable; approach problems in conventional, practical, and reasonable way; lack self-confidence; sincere, quiet, modest, withdrawn, seclusive; unpopular; overcontrolled; unlikely to express feelings openly.

Scale 10 (Social introversion)

High-scoring people (> 65) show: socially introversion more comfortable alone or with a few close friends; reserved, shy, and retiring; uncomfortable around members of opposite sex; hard to get to know; sensitive to what others think; troubled by lack of involvement with other people; overcontrolled; not likely to display feelings openly; submissive and compliant; overly accepting of authority; serious, slow personal tempo; reliable, dependable; cautious, conventional, unoriginal in approach to problems; rigid, inflexible in attitudes and opinions; difficulty making even minor decisions; enjoys work; gain pleasure from productive personal achievement; tend to worry; are irritable and anxious; moody, experience guilt feelings; have episodes of depression or low mood.

Low (T < 45): sociable and extroverted; outgoing, gregarious, friendly and talkative; strong need to be around other people; mix well; intelligent, expressive, verbally fluent; active, energetic, vigorous; interested in status, power and recognition; seeks out competitive situations; have problem with impulse control; act without considering the consequences of actions; immature, self-indulgent; superficial, insincere relationships; manipulative, opportunistic; arouses resentment and hostility in others.

Adapted from Butcher (1989).

In addition, prior to publication of the MMPI-2 there were a number of validity studies conducted on the revised form. For example, the MMPI revision committee collected personality ratings on more than 800 couples included in the normative sample. These personality ratings clearly cross-validated a number of the original scales. Moreover, validation research was conducted on a number of samples including schizophrenics and depressives (Ben-Porath, Butcher, & Graham, 1991); marital problem families (Hjemboe & Butcher, 1991); potential child-abusing parents (Egeland, Erickson, Butcher, & Ben-Porath, 1991); alcoholics (Weed, Butcher, Ben-Porath, & McKenna, 1992); airline pilot applicants (Butcher, 1994); military personnel (Butcher, Jeffrey et al., 1990).

Since the MMPI-2 was published in 1989 a number of other validation studies have been published (Archer, Griffin, & Aiduk, 1995; Ben-Porath, McCully, & Almagor, 1993; Blake et al., 1992; Clark, 1996; Husband & Iguchi, 1995; Keller & Butcher, 1991; Khan, Welch, & Zillmer, 1993).

4.14.2.2 Basic Personality Inventory

The BPI (Jackson, 1989) was published as an alternative to the MMPI-2 for the global assessment of psychopathology. The key aims in developing the BPI were to produce a broad-band measure of psychological dysfunctioning as measured by the MMPI that was: (i) relatively short, (ii) incorporated modern principles of test construction, and (iii) showed

Table 3 Description of the MMPI-2 content scales.

1. Anxiety (ANX)
High scorers report general symptoms of anxiety including tension, somatic problems (i.e., heart pounding and shortness of breath), sleep difficulties, worries, and poor concentration. They fear losing their minds, find life a strain, and have difficulties making decisions. They appear to be readily aware of these symptoms and problems, are willing to admit to them.

2. Fears (FRS)
A high score indicates an individual with many specific fears. These specific fears can include blood; high places; money; animals such as snakes, mice, or spiders; leaving home; fire; storms and natural disasters; water; the dark; being indoors; and dirt.

3. Obsessiveness (OBS)
High scorers have tremendous difficulties making decisions and are likely to ruminate excessively about issues and problems, causing others to become impatient. Having to make changes distresses them, and they may report some compulsive behaviors such as counting or saving unimportant things. They are excessive worriers who frequently become overwhelmed by their own thoughts.

4. Depression (DEP)
High scorers on this scale show significant depression. They report feeling blue, uncertain about their future, and uninterested in their lives. They are likely to brood, be unhappy, cry easily, and feel hopeless and empty. They may report thoughts of suicide or wishes that they were dead. They may believe that they are condemned or have committed unpardonable sins. Other people may not be viewed as a source of support.

5. Health concerns (HEA)
Individuals with high scores report many physical symptoms across several body systems. Included are gastrointestinal symptoms (e.g., constipation, nausea and vomiting, stomach trouble), neurological problems (e.g., convulsions, dizzy and fainting spells, paralysis), sensory problems (e.g., poor hearing or eyesight), cardiovascular symptoms (e.g., heart or chest pains), skin problems, pain (e.g., headaches, neck pains), respiratory troubles (e.g., coughs, hay fever, or asthma). These individuals worry about their health and feel sicker than the average person.

6. Bizarre mentation (BIZ)
Psychotic thought processes characterize individuals high on the BIZ scale. They may report auditory, visual, or olfactory hallucinations and may recognize that their thoughts are strange and peculiar. Paranoid ideation (e.g., the belief that they are being plotted against or that someone is trying to poison them) may be reported as well. These individuals may feel that they have a special mission or powers.

7. Anger (ANG)
High scorers tend to have anger control problems. These individuals report being irritable, grouchy, impatient, hotheaded, annoyed, and stubborn. They sometimes feel like swearing or smashing things. They may lose self-control and report having been physically abusive towards people and objects.

8. Cynicism (CYN)
High scorers tend to show misanthropic beliefs. They expect hidden, negative motives behind the acts of others, for example, believing that most people are honest simply for fear of being caught. Other people are to be distrusted, for people use each other and are only friendly for selfish reasons. They likely hold negative attitudes about those close to them, including fellow workers, family, and friends.

9. Antisocial practices (ASP)
High scorers tend to show misanthropic attitudes like high scorers on the CYN scale. The high scorers on the ASP scale report problem behaviors during their school years and other antisocial practices like being in trouble with the law, stealing or shoplifting. They report sometimes enjoying the antics of criminals and believe that it is all right to get around the law, as long as it is not broken.

10. Type A (TPA)
High scorers report being hard-driving, fast-moving, and work-oriented individuals, who frequently become impatient, irritable, and annoyed. They do not like to wait or be interrupted. There is never enough time in a day for them to complete their tasks. They are direct and may be overbearing in their relationships with others.

11. Low self-esteem (LSE)
High scores on LSE characterize individuals with low opinions of themselves. They do not believe that they are liked by others or that they are important. They hold many negative attitudes about themselves including beliefs that they are unattractive, awkward, and clumsy, useless, and a burden to others. They certainly lack self-confidence, and find it hard to accept compliments from others. They may be overwhelmed by all the faults they see in themselves.

Table 3 (continued)

12. Social discomfort (SOD)
High scorers tend to be very uneasy around others, preferring to be by themselves. When in social situations, they are likely to sit alone, rather than joining in the group. They see themselves as shy and dislike parties and other group events.

13. Family problems (FAM)
High scorers tend to show considerable family discord. Their families are described as lacking in love, quarrelsome, and unpleasant. They even may report hating members of their families. Their childhood may be portrayed as abusive, and marriages seen as unhappy and lacking in affection.

14. Work interference (WRK)
High scorers tend to show behaviors or attitudes that are likely to contribute to poor work performance. Some of the problems relate to low self-confidence, concentration difficulties, obsessiveness, tension and pressure, and decision-making problems. Others suggest lack of family support for the career choice, personal questioning of career choice, and negative attitudes towards co-workers.

15. Negative treatment indicators (TRT)
High scorers tend to show negative attitudes towards doctors and mental health treatment. High scorers do not believe that anyone can understand or help them. They have issues or problems that they are not comfortable discussing with anyone. They may not want to change anything in their lives, nor do they feel that change is possible. They prefer giving up, rather than facing a crisis or difficulty.

Adapted from: Butcher, Graham, Williams, & Ben-Porath (1990).

empirical evidence of being able to discriminate between normal and dysfunctional persons as well as being able to predict pathological behavior.

The BPI is made up of 240 items, grouped into 12 × 20 item scales. Neurotic tendencies are measured through the scales of Hypochondriasis, Depression, Anxiety, Social Introversion, and Self-depreciation. Aspects of sociopathy are measured by Denial, Interpersonal Problems, Alienation, and Impulse Expression scales. Psychotic behavior is assessed by scales labeled Persecutory Ideas, Thinking Disorder, and to some degree, by the Deviation scale. The Deviation scale comprises 20 critical items that are intended to serve as the basis for further clinical follow up. In contrast, the definitions of the constructs reflected in the other 11 items on the BPI are based on the results of a multivariate analysis of the content underlying the MMPI and the Differential Personality Inventory (Jackson & Messick, 1971).

The strong internal psychometric properties of the BPI attest to its careful construction (Jackson, 1989). Item properties and item factor analyses support the internal structure of the instrument. Both internal consistencies and test–retest reliability estimate fall in the 0.70 to 0.80 range. Various validity studies (some published in the manual, others in the literature) show that the BPI can indeed discriminate between normal and non-normal (e.g., delinquent) persons and can, within psychiatric populations, predict a variety of clinical criteria. Norms exist for adolescents and adults, spanning a variety of sample types such as community, psychiatric, college, and forensic. However, these norms are almost entirely based on white populations. Moreover, most of the normative sample was collected using nonstandard data collection procedures. For example, booklets were mailed to subjects instead of being administered under standard conditions. In addition, each of the subjects in the normative sample responded to one-third of the items in the booklet, an artifact that makes it difficult to perform some analyses on the normative sample (e.g., alpha coefficients).

Among the difficulties that have been associated with the BPI is a lack of validity scales for identifying invalid response protocols. The one content scale that would logically appear to have some bearing on the issue of protocol validity is the Denial scale, which is described in the manual as a measure of lack of insight and lack of normal affect. Unfortunately, Denial appears to be a relatively weak scale in terms of its reliability and validity in various empirical studies (Holden et al., 1988; Jackson, 1989). Two other complaints that have been voiced are a lack of an established link to diagnostic categories and a lack of work on profile interpretation.

Overall, the BPI has shown psychometric potential as a general measure of psychopathology. The basic developmental work on it is sound. What the BPI now needs are more extensive norms along with further work on its clinical applicability.

4.14.2.3 Personality Assessment Inventory

The PAI (Morey, 1991) is another inventory of general psychopathology. The PAI is a 344

item self-report measured designed to screen for approximately the same pathological domains as the MMPI/MMPI-2. It is used to collect information related to diagnosis, and to provide input on treatment planning.

The PAI includes four validity scales, 11 clinical scales, five treatment scales and two interpersonal scales. The clinical scales are Somatic Complaints, Anxiety-related Disorders, Depression, Mania, Paranoia, Schizophrenia, Borderline Features, Antisocial Features, Alcohol Problems, and Drug Problems. Treatment scales are Aggression, Suicidal Ideation, Stress, Nonsupport, and Treatment Rejection. Interpersonal scales are Dominance and Warmth. Of these scales, 10 are further divided into subscales that are intended to measure distinct constructs. A total of 27 items are designated critical items which, according to the author, should be followed up. This does seem to be a relatively large number of scales to interpret meaningfully in view of the total number of items.

Evidence for the reliability of the 22 scales is generally good. Across normative, clinical and college samples, median alpha coefficients are all in the 0.80–0.90 range. One month test–retest coefficients are reported to be in the 0.80s in the manual and in the 0.70s in the literature (Boyle & Lennon, 1994). Interestingly, some of the lowest reliabilities are found for the validity scales, a phenomenon that has been found for the validity scales on other instruments such as the MMPI (e.g., Fekken & Holden, 1991) and may well be a function of range restriction in the scores on such scales.

Normative PAI data for the USA are extensive. Standardization samples for a census-matched group, a clinical sample representing 69 sites, and a college sample drawn from seven different US universities each number over 1000 respondents. Normative data in the manual are also reported separately by age, education, gender, and race. Normative information for other countries (Canada, Australia, the UK, etc.) would be desirable.

In view of the recency of the publication of the PAI, there are only a handful of studies that bear on its validity. The manual reports evidence of the concurrent validity of the PAI in the form of correlations with other measures of psychopathology. A number of other studies show that the PAI can discriminate between diagnostic groups. Boyle and Lennon (1994) showed that the PAI can distinguish normals, alcoholics, and schizophrenics. Schinka (1995) was further able to use the PAI to develop a typology for alcoholics that had validity with several external variables. Finally, Alterman and colleagues (1995) demonstrated that methadone maintenance patients scored differently from the normative populations.

Generally, preliminary data suggest that the PAI is a well-constructed and brief measure of psychopathology. More research on its clinical validity needs to be completed before it can be considered to be a useful measure of psychopathology in clinical settings. There are no data to support the PAI's use in lieu of the MMPI-2 which has a more substantial empirical database.

4.14.3 SPECIALIZED OR FOCUSED CLINICAL ASSESSMENT MEASURES

In this section we will address several other measures that have been developed for clinical assessment and research to assess specific or more narrowly focused characteristics rather than omnibus instruments such as the MMPI-2. Several of these instruments will be examined to illustrate their application and potential utility for evaluating specific problems in clinical settings.

4.14.3.1 Millon Clinical Multiaxial Inventory

The MCMI (Millon, 1977, 1987, 1994) was developed by Theodore Millon for making clinical diagnoses on patients. The MCMI was intended to improve upon the long-established MMPI. In contrast to the MMPI/MMPI-2, the MCMI was designed with fewer items; is based on an elaborate theory of personality and psychopathology; and explicitly focuses on diagnostic links to criteria from the *Diagnostic and statistical manual of mental disorders* (*DSM*).

The MCMI was developed rationally rather than empirically. Millon has stated in his three test manuals (1977, 1987, 1994) as well as elsewhere (Millon & Davis, 1995) that development of the MCMI is to be an ongoing process. To keep the MCMI maximally useful for clinical diagnosis and interpretation, it must be continually updated in view of theoretical refinements, empirical validation studies, and evolutions in the official *DSM* classification systems. Most updated test manuals leave the test user with the impression that the developer considered test revision a necessary evil. Very rarely do you see continuous improvement as a test developer's goal, in part because this ever-changing process makes the accumulation of a solid research base difficult.

All three MCMI versions comprise 175 true/false items. However, across versions, the exact test items have evolved through revision or replacement. The number of scales and validity indices that can be calculated from these items has also increased. The original version, the MCMI-I, had 20 clinical scales and two validity scales. The MCMI-II yielded 22 clinical scales and three validity scales. The current MCMI-III has 24 clinical scales, three modifying indices and a validity index. Many items appear on several scales making for great item overlap.

On the MCMI-III, 14 clinical scales assess personality patterns that relate to *DSM-IV* Axis II disorders. Another 10 scales measure clinical syndromes related to *DSM-IV* Axis I disorders. The modifying indices, Disclosure, Desirability, and Debasement, are correction factors applied to clinical scale scores to ameliorate respondents' tendencies to distort their responses. The validity index comprises four bizarre or highly improbable items meant to detect careless, random, or confused responding.

The relationship between scales and items is explicated in detail in the manual. Millon started with a theory-based approach to writing items, followed by an evaluation of the internal structure of the items, and finally engaged in an assessment of the diagnostic efficiency of each item for distinguishing among diagnostic groups before final placement of an item on a scale. Millon departed from usual psychometric practice in a way that results in some unfortunate complications. Item overlap across scales is permitted: on average, items appear on three different scales with differential weights. This makes scoring inordinately complex, which makes assessment of scale homogeneity complex, and in turn this makes evaluation of the empirical structure underlying the scales complex. There are technical solutions for these problems but the result is that the MCMI is not an easy instrument with which to work.

Despite its psychometric drawbacks, the theory underlying the MCMI is generally agreed to be elegant and a substantial asset (McCabe, 1987; Reynolds, 1992). Each dimension measured by the test has a clear conceptual link to Millon's theory of psychopathology. Such a theory allows for the generation of clinical inferences based on a small number of fundamental principles (Millon & Davis, 1995). Not only do these inferences guide measurement, but also they enhance understanding of the constructs, bear on practical treatment decision, and produce research hypotheses.

One of the stated goals of the MCMI is to place patients into target diagnostic groups. To this end, the MCMI scales are directly coordinated with the *DSM* diagnostic categories. How well does the MCMI live up to its aim? The manuals report good evidence of diagnostic efficiency. However, the recent literature (available on the MCMI-I and MCMI-II) suggests the following three generalizations. First, the MCMI has only modest accuracy for assigning patients to diagnostic groups across a variety of clinical criteria (e.g., Chick, Martin, Nevels, & Cotton, 1994; Chick, Sheaffer, Goggin, & Sison, 1993; Flynn, 1995; Hills, 1995; Inch & Crossley, 1993; Patrick, 1993; Soldz, Budman, Demby, & Merry, 1993). Second, the MCMI may be better at predicting the absence than the presence of a disorder (Chick et al., 1993; Hills, 1995; Soldz et al., 1993). Third, the MCMI may be better at predicting some types of disorders than others but there is little agreement on which ones (Inch & Crossley, 1993; Soldz et al., 1993).

One source of the difficulty may be the base rate scores. Raw scores on scales are weighted and converted to base rate scores. The base rate scores reflect the prevalence of a particular personality disorder or pathological characteristic in the overall population. Their use is intended to maximize the number of correct classifications relative to the number of incorrect classifications when using the MCMI to make diagnoses (Millon & Davis, 1995). If the estimated base rates for the various diagnostic categories are poor, then the predictive accuracy of the MCMI can be expected to be poor (Reynolds, 1992).

One negative consequence of the type of norms used in the development of the MCMI inventories is that they do not discriminate between patients and normals. Use of the MCMI assumes that the subject is a psychiatric patient. Consequently, the MCMI overpathologizes individuals who are not actually patients. The MCMI should not be used where issues of normality need to be addressed. For example, if the test were used in family custody evaluations or personnel screening, the test interpretation would appear very pathological—it cannot do otherwise.

How does the MCMI compare to the MMPI? The MCMI publisher appears to emphasize that the MCMI and MMPI-2 measure different characteristics and the MCMI is shorter to administer to patients. Whereas the MMPI measures a broad range of psychopathology, the MCMI has its premier focus on the assessment of personality disorders. Consonant with its rational construction, the elaborate theoretical underpinnings of the MCMI are impressive. In contrast, however, the test literature that supports the validity of the MMPI/MMPI-2 is not available for the MCMI. Validation research on it has not proceeded at a very high pace. Whether the MCMI will have either the clinical utility or the heuristic value that the MMPI enjoys remains unanswered until more clinical research, and perhaps more refinements, are undertaken with the MCMI.

4.14.3.2 The Beck Depression Inventory

The BDI was first introduced in 1961, and it has been revised several times since (Beck et al., 1988). The BDI has been widely used as an assessment instrument in gauging the intensity of depression in patients who meet clinical diagnostic criteria for depressive syndromes. However, the BDI has also found a place in research with normal populations, where the focus of use has been on detecting depression or depressive ideation.

The BDI was developed in a manner similar to the MMPI: clinical observations of symptoms and attitudes among depressed patients were contrasted to those among nondepressed patients in order to obtain differentiation of the depressed group from the rest of the psychiatric patients. The 21 symptoms and attitudes contained in the BDI reflect the intensity of the depression; items receive a rating of zero to three to reflect their intensity and are summed linearly to create a score which ranges from 0 to 63. The 21 items included reflect a variety of symptoms and attitudes commonly found among clinically depressed individuals (e.g., Mood, Self-dislike, Social Withdrawal, Sleep Disturbance). The BDI administration is straightforward, and it can be given as an interview by the clinician or as a self-report instrument (requiring a fifth or sixth grade reading level).

The BDI is interpreted through the use of cut-off scores. Cut-off scores may be derived based on the use of the instrument (i.e., if a clinician wishes to identify very severe depression, then the cut-off score would be set high). According to Beck et al. (1988), the Center for Cognitive Therapy has set the following guidelines for BDI cut-off scores to be used with affective disorder patients: scores from 0 through 9 indicate no or minimal depression; scores from 10 through 18 indicate mild to moderate depression; scores from 19 through 29 indicate moderate to severe depression; and scores from 30 through 63 indicate severe depression.

Two important issues must be considered by clinicians regarding the results of the BDI. Unlike the MMPI/MMPI-2 and other major self-report instruments, the BDI has no safeguards against faking, lying, or variable response sets. Thus, clinicians are warned against this drawback of the BDI in assessing depressive thoughts and symptoms. In settings where faking or defensiveness are probable threats to the validity of the test, clinicians may need to reconsider their use of the BDI. The other issue pertains to the state–trait debate in assessment. The BDI is extremely sensitive to differences in the instructions given to an examinee such that certain instructions yield a state-like index of depressive thinking, whereas, other instructions yield a more trait-like index of depressive thinking. Again, clinicians are encouraged to use caution when administering the BDI and to tailor the administration instructions to the type of index (state or trait) that is desired.

In their review of the psychometric properties of the BDI, Beck et al. (1988) reported high internal consistency reliability of the instrument among both psychiatric and nonpsychiatric populations. The authors also reported that the BDI closely parallels the changes in both patient self-report and clinicians' ratings of depression (i.e., the BDI score accurately reflects changes in depressive thinking). Finally, they also presented evidence for the content, concurrent, discriminant, construct, and factorial validity of the BDI.

The acceptable reliability and validity of the BDI have helped make it a widely used objective index of depressive thinking among clinicians. Perhaps the most obvious use of the BDI is as an index of change in the level or intensity of depression. With an increasing focus on managed healthcare and accountability by psychotherapeutic service providers, the BDI offers a reliable and valid index of depressive symptoms and attitudes which can be used effectively to document changes brought about in therapy.

4.14.3.3 The State-Trait Anxiety Inventory

The STAI was developed by Spielberger et al. (1970) to measure anxiety from the perspective of states vs. traits. The state measurement assesses how the individual feels "right now" or at this moment. Subjects are asked to rate the intensity of their anxious feelings on a four point scale as to their experience of feelings in terms of: not at all, somewhat, moderately so, or very much so. The trait anxiety measure addresses how the individuals generally feel by rating themselves on a four-point scale: almost never, sometimes, often, or almost always.

Since it was developed in 1966 the STAI has been translated into over 48 different languages and has been widely researched in a variety of clinical and school settings (Spielberger, Ritterband, Sydeman, Reheiser, & Unger, 1995). The evidence for construct validity of the STAI comes from a variety of sources, for example, correlations with other anxiety measures (Spielberger, 1977), clinical settings (Spielberger, 1983), and medical and surgical patients (Spielberger, 1976). Nonetheless, there has been general debate in the literature about the conceptual and practical benefits of the trait vs. state distinction.

The STAI has become a very widely used measure in personality and psychopathology research in the USA and in other countries. However, the state-trait inventories have not been as broadly used as clinical assessment instruments.

4.14.3.4 Whitaker Index of Schizophrenic Thinking

The WIST (Whitaker, 1973, 1980) was developed to measure the type of thought impairment that differentiates between schizophrenic and "normal" thought processes. The WIST is intended to be individually administered as a screening tool or as one part of a battery of tests. Its multiple choice format makes the WIST a test that is easy to give and to score.

On the WIST, schizophrenic thought is defined as a discrepancy between actual and potential performance on cognitive reasoning tasks. This impaired thinking has three components: (i) a degree of illogicality, as reflected in inappropriate associations and false premises; (ii) a degree of impairment relative to previous performance, as reflected in slowness; and (iii) a degree of unwittingness, as reflected in unawareness of the incorrectness of responses. Whitaker's definition of schizophrenic thought is carefully explicated in the manual, making it possible for the test user to understand exactly what the WIST is measuring. Whitaker, however, has been criticized for basing his definition on a narrow reading of the literature on schizophrenic thought disorder (Payne, 1978).

The WIST itself is made up of 25 multiple choice items that are divided into three subtests: Similarities, Word Pairs, and New Inventions. Each item consists of a stimulus and five response options that differ in degree of illogicality. To illustrate, consider this sample item for the Similarities subtest: car: automobile, tires, my transportation, jar, smickle. The correct answer receives a score of 0; a loose association, reference idea, clang association, and nonsense association would receive scores of 1, 2, 3, and 4, respectively. The test administrator presents for a second time any items that the respondent answered incorrectly on the initial test taking.

Three scores are calculated for the WIST: total Score for all response alternatives selected either in the original or in the second enquiry phase; total Time for the initial completion of the WIST; and an overall Index that combines WIST Score and WIST Time. Presumably the WIST Score and Time components are added together because they both relate to schizophrenic thought disorder. However, a review of the empirical data supporting this conceptualization finds mixed results at best (Fekken, 1985). One view is that the predictive efficiency of the WIST is not enhanced by including the Time component.

There are two parallel forms of the WIST which differ in content. Items on Form A are intended to be stressful or anxiety provoking. They assess oral-dependency, hostility, and manifestly sexual content. The content of Form B is neutral. There has been little formal evaluation of the comparability of the two forms. Although some studies have reported similar rates of diagnostic efficiency for the two forms (Evans & Dinning, 1980; Leslie, Landmark, & Whitaker, 1984), overall validity evidence would suggest that Form A is stronger than Form B. Additional work to clarify the comparability of Forms A and B would have implications for selecting which form to use and for using these alternate WIST forms to assess changes in symptomatology.

There is relatively little information on reliability available on the WIST. The manual does report intratest reliabilities of the two WIST forms as Hoyt's reliability coefficients of around 0.80. Test–retest reliability data, either on the WIST subscores or subscales, are not provided in the manual nor do they appear to be readily available in the literature. Similarly, alternate forms of reliability, calculated as correlations between subscales, or subscores, are not readily available.

There is reasonable evidence for the convergent validity of the WIST. Two reviews report that WIST scores tend to have a 60–70% agreement with systems for diagnosing schizophrenia (Fekken, 1985; Grigoriadis, 1993). The WIST has been empirically associated with other indices of schizophrenia such as the MMPI/MMPI-2 Sc scale (Evans & Dinning, 1980; Fishkin, Lovallo, & Pishkin, 1977; Grigoriadis, 1993) although not with other schizophrenia indices including conceptually relevant SCL-90 scales (Dinning & Evans, 1977) and the New Haven Schizophrenic Index (Knight, Epstein, & Zielony, 1980).

Discriminant validity of the WIST has been harder to establish. The WIST has difficulty distinguishing among different psychiatric diagnostic groups (e.g., Burch, 1995; Pishkin, Lovallo, & Bourne, 1986). This is a serious shortcoming because the WIST is likely to be used in clinical settings precisely to help make such distinctions. A second problem with discriminant validity has been the tendency of the WIST to correlate negatively with measures of general cognitive ability. Based on such data one could share the view of at least one pessimistic reviewer and claim that the WIST has no demonstrated use (Payne, 1978).

Alternatively, the WIST may have a role in a comprehensive assessment battery. The WIST has never been promoted as a stand-alone test, nor has it ever been promoted as a comprehensive measure of the full range of schizophrenic symptomatology. Rather, the data on the WIST may be thought of as a general measure of cognitive deficit rather than cognitive deficit specific to schizophrenia. Thus, it may provide one objective source of data for accepting or rejecting a more general diagnosis of psychosis.

4.14.4 NORMAL RANGE PERSONALITY ASSESSMENT

Identification and description of psychological disorder is but one reason to administer a personality measure to an adult. Objective personality measures also are widely used outside the clinic or hospital setting for normal range assessment situations. Several objective personality measures will be discussed regarding their use in research, educational, vocational assessment, and personnel selection.

4.14.4.1 Objective Personality Measures in Research

4.14.4.1.1 The Five Factor Model (the Big Five)

One of the most popular current conceptualizations being studied today involves taxonomies of personality traits based on factor analytic methods and is commonly referred to as the "Big Five," or the Five Factor Model (FFM). As Butcher and Rouse (1996) note in their review, some personality researchers have rejected the FFM as an end-all to personality trait theory, while others in the area of personality research continue to embrace it. Among the proponents of the FFM are Costa and his colleagues, who have proposed the NEO Personality Inventory (NEO-PI) as a self-report measure of the FFM personality dimensions (Costa & McCrae, 1992). The FFM consists of five factor-analytically derived dimensions of personality. The five dimensions (are Neuroticism (N), Extroversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C). Neuroticism has long been a familiar adjective among clinicians for describing people who tend to experience psychological distress. At the opposite end of the Neuroticism dimension is Emotional Stability, which represents the tendency to stay on a psychologically even keel. Extroversion encompasses the concepts of positive emotionality, sociability, and activity. The Extroversion dimension is rather broad in its scope, although lay people often use the term narrowly to refer to general "outgoingness." Openness to Experience represents, broadly, a person's level of constriction in their experiencing of the world; it is often associated with creativity (and even hypnotizability). Agreeableness represents the dimension of interpersonal behavior; the counterpart of Agreeableness is Antagonism. Finally, Conscientiousness represents a dimension of scrupulous organization of behavior.

4.14.4.1.2 NEO Personality Inventory

The NEO-PI is a 181-item inventory designed to index the FFM personality dimensions (N, E, O, A, and C); the NEO-PI also yields several subscales of the N, E, and O dimensions (Costa & McCrae, 1992). There is a self-report form of the NEO-PI as well as an observer-rating form. Although the authors of the NEO-PI argue for its utility in clinical settings, the instrument has been studied almost exclusively in nonclinical populations. One example of a recent investigation using the NEO-PI in research with psychiatric samples is the examination of differences in stimulus intensity modulation among older depressed individuals with and without mania features (Allard & Mishara, 1995). The NEO-PI was used to examine the hypothesis that stimulus intensity augmenters with unipolar depression would be introverted, whereas reducers with bipolar depression would be extroverted. Determinations of depression with and without mania features were based on scores on the MMPI. Costa and McCrae (1992) point to the NEO-PI as a useful assessment tool in aiding the clinician with understanding the client, selection of treatment, and even anticipating the course of therapy. However, the NEO-PI has not found wide use among clinicians to date (Butcher & Rouse, 1996).

Unfortunately, the NEO-PI is very susceptible to faking (Bailley & Ross, 1996) and does not contain validity indices to detect deviant response sets.

4.14.4.1.3 Multidimensional Personality Questionnaire

The Multidimensional Personality Questionnaire (MPQ) is a 300-item self-report instrument that was developed by Tellegen (unpublished manuscript) in an attempt to clarify the "self-view domain" in personality research. Tellegen used a classical iterative test construction approach involving several rounds of factor analysis to come up with the 11 primary scales, six validity scales, and three "higher-order" factors. The 11 primary scales (which include the dimensions of Social Potency, Control, Harm Avoidance, Well-being, Aggression, and others) load onto the three higher-order scales, which represent the familiar personality domains of Positive Affectivity, Negative Affectivity, and Constraint. The six validity indices include VRIN and TRIN, which are conceptually similar to those on the MMPI-2.

The MPQ is not as widely used as other objective personality measures in either the clinical or the research domain. However, recent investigations suggest a place for the MPQ in normal range and clinical personality assessment. For example, Kuhne, Orr, and Baraga (1993) demonstrated the utility of certain MPQ scales for discriminating among veterans with and without post-traumatic stress disorder. Krueger et al. (1994) administered the MPQ to adolescents in their community-based longitudinal study and found that certain MPQ scales were useful in distinguishing those who engaged in delinquency from those who abstained. These two reports suggest the utility of the MPQ both in clinical settings and in normal range personality assessment.

4.14.4.1.4 Personality Research Form

One well-known measure of normal personality is the Personality Research Form (PRF). Developed by Douglas N. Jackson (1984), the PRF is a true/false, multiscale measure of 20 of the psychosocial needs (e.g., achievement, aggression, sociability) originally defined by Henry Murray (1938). Many psychometrics texts hold up the PRF as a model of the construct approach to test construction. Indeed, much of the appeal of the PRF lies in its ability to measure a large number of normal personality characteristics while minimizing both scale intercorrelations and the influence of social desirability and acquiescence. Moreover, a variety of validity studies have been published in the 1980s and 1990s supporting the psychometric soundness of the PRF. Critics of the PRF complain that despite its technical elegance the PRF does not reflect an integrated model of personality, which limits its applicability to real-life testing situations. Recent research, however, shows that the content assessed by the PRF may be well described by the Five Factor structure (Paunomen, Jackson, Trzebinski, & Fosterling, 1992). In research the popularity of the PRF remains high as attested to by the number of references in bibliographies, such as the one produced by MacLennan in 1991, which lists over 375 studies featuring the PRF in the literature.

4.14.4.1.5 Sixteen Personality Factor Test

The 16PF was originally developed in the 1940s by Raymond Cattell to measure the primary factors of normal personality. At that time, Cattell's unique contribution was to apply factor analysis as a method for uncovering the full scope of personality. The fifth and current edition of the 16PF (Cattell, Cattell, & Cattell, 1993) measures the well-known 16 personality factors, plus it summarizes these factors into five global factors which again bear similarity to the well-known "Big Five." The global factors are: extroversion, anxiety, tough-mindedness, independence, and self-control. Relative to earlier editions, the fifth editions includes updated language; fewer items; and improved reliability and response style scales to measure impression management, infrequent responding, and acquiescence. Because of the recency of the publication of the fifth edition, there are few studies pertaining to its validity. However, a large database supports the validity of earlier editions of the 16PF. The 16PF has applicability in clinical, educational, organizational, and research settings. With the expansion of its interpretive reports and profiles, the 16PF would appear to be particularly useful in personal or vocational counseling settings.

4.14.4.1.6 California Psychological Inventory

The CPI developed by Gough (1957) is a multiscale, objective, self-report instrument used with normal range and psychiatric populations (Megargee, 1972). The CPI is similar in structure and content to the MMPI (and the MMPI-2). In fact, many of the items on the CPI are identical in wording to the items on the MMPI. The CPI focuses on "everyday" concepts about personality, such as dominance and responsibility among others. Eighteen scales (divided into four classes) are derived from the CPI; the results of the test are interpreted by reference to the plotted profile of standard scores. The body of literature using the CPI in normal range assessment is quite extensive in both quantity and breadth. Investigations include the use of the CPI to identify personality types, based on profiles, among a group of college students (Burger and Cross, 1979) and an examination of the underlying personality structure as measured by the CPI, again using a sample of college students (Deniston and Ramanaiah, 1993). The CPI also has been employed in the assessment of psychiatric samples. Especially noteworthy is the utility of the CPI with criminal samples (see Laufer, Skoog, and Day, 1982, for a relevant review of this literature).

Some specific applications of psychological tests in "normal range" settings are now described.

4.14.4.2 Objective Personality Measures in Educational/Vocational Assessment

The use of objective personality measures has become increasingly popular among professionals in the field of educational/vocational assessment. Research has identified the FFM, the NEO-PI, and the CPI as particularly useful in educational/vocational assessment.

4.14.4.2.1 FFM

The Big Five personality dimensions of the FFM have been studied in several contexts related to educational/vocational assessment. Moreover, because the FFM is a theoretical concept, various instruments have been used to assess the relationship between the Big Five personality dimensions and various educational/vocational assessment variables. Two recent investigations illustrate the utility of the Big Five personality dimensions in identifying candidates for admission to educational institutions and identifying characteristics among students in various university programs. Williams, Munick, Saiz, and Formy-Duval (1995) found that a mock graduate school admissions board (composed of graduate school faculty) favored for admission those hypothetical candidates whose applications reflected high ratings on the Big Five dimensions of Conscientiousness and Openness to Experience. Conversely, the Big Five dimensions of Extroversion and Agreeableness were not associated with a favorable impression of hypothetical candidates. Kline and Lapham (1992) assessed the Big Five personality dimensions among a group of college students to examine personality differences among the various fields of study (i.e., between different college majors). Students of various majors were not discriminated by levels of either Neuroticism or Extroversion. However, students in two fields of study (science and engineering) were marked by high ratings on the Big Five personality dimensions of Conscientiousness and Conventionality.

4.14.4.2.2 NEO-PI

The personality dimensions that underlie the NEO-PI appear to be related to at least one well-known typology of vocational personalities (which, in turn, correspond to vocational preferences). Gottfredson, Jones, and Holland (1993) found a relationship between the Big Five personality dimensions assessed by the NEO-PI and the six vocational personality dimensions proposed by Holland (assessed with the Vocational Preferences Inventory). In general, there was overlap among two to four of the significant factors extracted from each of the assessment instruments. However, the NEO-PI Neuroticism, Likability, and Control factors were not represented in the Holland vocational personality dimensions, which suggests a distinctive and qualitatively different role for objective personality assessment in vocational counseling. Other work with the NEO-PI has shown that two of the Big Five personality dimensions, Neuroticism and Agreeableness, are strongly related to occupational burnout among healthcare workers (Piedmont, 1993). Healthcare workers with higher ratings on the Neuroticism dimension were more likely to experience occupational burnout. Conversely, workers with higher ratings on the Agreeableness dimension were less likely to succumb to occupational burnout.

4.14.4.2.3 CPI

The CPI is one of the most widely used personality instruments in normal range assessment in educational/vocational settings. Walsh (1974) examined personality traits among college students identified as making hypothetical career choices that were either congruent or incongruent with their vocational personality type (as proposed by Holland). It was found that congruent students could be described by their CPI profiles as socially accepted, confident, and planful, whereas incongruent students could be described as impulsive, unambitious, and insecure. The well-known Strong Vocational Interest Blank (SVIB) used commonly among vocational and educational counselors is related to various personality traits as well. Johnson, Flammer and Nelson (1975) found a relationship among SVIB factors and CPI personality factors, especially those related to the global introversion/extroversion personality dimension.

4.14.4.2.4 MMPI/MMPI-2

Although it was not designed for educational research or placement purposes the MMPI/MMPI-2 has been among the most frequently employed instruments in this context. In a recent survey of test use in personnel and educational screening the MMPI has been employed effectively in a number of studies, for example: Anderson (1949), Appleby and Haner (1956), Applezweig (1953), Barger and Hall (1964), Barthol and Kirk (1956), Burgess (1956), Centi (1962), Clark (1953, 1964), and Frick (1955) to mention only a few.

It appears evident that personality assessment contributes information independent of direct educational/vocational interest measures. There is extensive information available on the use of the MMPI/MMPI-2 in this setting. As demonstrated in the research with the NEO-PI and the CPI, objective personality assessment instruments yield information relevant to vocational assessment over and above that given by narrowly defined vocational personality assessments (like the one proposed by Holland). Moreover, elements of the FFM and independent personality constructs of the CPI have proven their utility in educational assessment in diverse areas such as admissions to educational institutions and choice of major fields of study. Clearly, objective personality assessment has carved out a valuable niche in normal range educational/vocational assessment.

4.14.4.3 Personnel Screening

Among the earliest and most extensive uses of personality tests with normals has been for the purpose of employment screening. The first formal North American, English language personality inventory, the Woodworth Personnel Data Sheet, was developed to screen out unfit draftees during World War I. A number of other personality questionnaires were developed in the 1930s to aid in personnel selection decisions.

The development and use of the MMPI during World War II provided a means for assessment psychologists to detect psychological problems that might make people unsuitable for key military assignments. Early research on the use of the MMPI centered around pilot selection and selection of nuclear submarine crewman. Following World War II, the MMPI came to be widely used in personnel selection particularly for occupations that required great responsibility or involved high stress such as air flight crews (Butcher, 1994; Cerf, 1947; Fulkerson, Freud, & Raynor, 1958; Fulkerson & Sells, 1958; Garetz & Tierney, 1962; Geist & Boyd, 1980; Goorney, 1970; Jennings, 1948); police and other law enforcement personnel (Beutler, Nussbaum, & Meredith, 1988; Beutler, Storm, Kirkish, Scogin, & Gaines, 1985; Butcher, 1991; Dyer, Sajwaj, & Ford, 1993; Hargrave & Hiatt, 1987; Saxe & Reiser, 1976; Scogin, & Beutler, 1986; Scogin & Reiser, 1976); and nuclear power employees (Lavin, Chardos, Ford & McGee, 1987).

Following the revision of the MMPI and publication of the MMPI-2, the revised instrument has been used in personnel screening (Butcher & Rouse, 1996) and currently is the most frequently used personality measure in personnel screening situations, particularly when the position is one that requires good mental health, emotional stability, and responsible behavior.

The MMPI-2 is usually employed in personnel selection to screen out candidates who are likely to have psychological problems from critical occupations such as police officer, airline pilot, and nuclear power control rooms, fire department, or air traffic control personnel.

4.14.4.4 Other Personality Measures in Personnel Selection

Objective measures of personality characteristics have a role in personnel selection similar to that in educational/vocational assessment. Namely, professionals in the field of personnel selection are interested in knowing which personality variables aid in the selection of quality employees. The literature suggests a valuable role for objective personality measures in the normal range assessment field of personnel selection.

4.14.4.5 The 16PF

One of the most widely used personality scales for employment screening is the 16PF. This inventory, with broad-ranging employment-relevant personality items, has been widely used in different contexts including: law enforcement (Burbeck & Furnham, 1985; Fabricatore, Azen, Schoentgen, & Snibbe, 1978; Hartman, 1987; Lawrence, 1984; Lorr & Strack, 1994; Topp & Kardash, 1986); pilots (Cooper & Green, 1976; Lardent, 1991); cabin crew personnel (Furnham, 1991); managers (Bartram, 1992; Bush, & Lucas, 1988; Chakrabarti, & Kundu, 1984; Henney, 1975); occupational therapists (Bailey, 1988); church counselors (Cerling, 1983) and teachers (Ferris, Bergin, & Wayne, 1988). The 16PF provides information about personality functioning and is typically employed to screen employees for positive personality features.

4.14.4.6 FFM

Barrick and Mount (1991) conducted a meta-analysis of the Big Five personality dimensions and their relationship to three criterion variables (job proficiency, training proficiency, and personnel data) within five occupational groups (professionals, police, managers, skilled/semiskilled and sales). Conscientiousness was related to each of the five occupational

groups. Additionally, Conscientiousness was related to the three criterion variables. These findings lead the authors to conclude that Conscientiousness is a personality trait related to job performance across occupational types. In a similar vein, Dunn, Mount, Barrick, and Ones (1995) found that the Conscientiousness personality dimension was related to managers' ratings of applicant hireability. This held true for various job types (medical technologist, carpenter, secretary, etc.). Taken together, these two reports suggest that professionals in the field of personnel selection can gain valuable information from objective personality measures of the FFM. Specifically, the Conscientiousness dimension, when taken with other relevant application information, may be a valuable discriminator among prospective job applicants.

4.14.4.7 NEO-PI

While research suggests that the Big Five personality dimensions add qualitative information to the personnel selection process, it may be that the various instruments used to measure the Big Five have differential validity across populations. Schmit and Ryan (1993) administered a shortened version of the NEO-PI to a sample of college students and to a sample of government job applicants. The FFM structure fitted the student population better than it fitted the job applicant population, which suggests a note of caution in putting too much weight on the Big Five personality dimensions in personnel selection. The authors note that job applicants are under different situational demands, which may affect their approach to a personality questionnaire (e.g., they may adopt a defensive response style). The lack of validity scales to assess response styles clearly limits the NEO-PI for this application. A recent study by Bailley and Ross (1996) showed that the NEO-PI is quite vulnerable to faking and is limited in not having scales to detect deviant response attitudes.

4.14.4.7.1 CPI

The NEO-PI and other personality measures have been examined in personnel selection studies covering a wide range of occupational groups (managers, secretaries, carpenters, etc.). However, the CPI has a long tradition of use in the selection of a specific occupation: law enforcement officers. Hargrave and Hiatt (1989) examined the ability of the CPI scales to differentiate police cadets rated by their instructors as suitable or unsuitable for the job of law enforcement officer. Several scales (including Sense of Well-being, Sociability, and Social Presence) differentiated suitable from unsuitable cadets. In the second study of the report, the authors compared incumbent law enforcement officers who either had or had not experienced serious on-the-job problems (e.g., providing drugs to inmates, excessive use of force). The Socialization scale was among the best discriminators between groups. The authors concluded that the CPI is a useful aid to the selection of law enforcement officers.

In addition to the literature on law enforcement screening, the CPI may be useful as an index of work performance in other fields. Toward this end, Hoffman and Davis (1995) validated the Work Orientation and Managerial Potential scales for the CPI on groups of employees in an entertainment facility. However, several of the original CPI scales performed as well as the two new scales in predicting job performance, which questions the need for additional CPI scales in the selection of personnel.

In summary, the use of objective personality measures in the selection of job applicants has proven a worthwhile endeavor overall. The Big Five personality dimensions, especially Conscientiousness, might be useful to the personnel selection process especially in the assessment of conscientiousness. Professionals involved in the screening of law enforcement candidates can use the CPI to differentiate suitable from unsuitable candidates. The CPI may also be useful in discriminating among groups of job applicants (although its primary use has been in the law enforcement field).

The scales on the 16PF have been shown to be relevant to personality descriptions that are useful in personnel selection. Finally, when it comes to evaluating potential psychopathology in potential employees the MMPI-2 is usually the instrument most employed.

Although objective personality measures appear to have a place in personnel selection, all information on the candidate must be weighed, as several personality researchers have noted the problem of response sets associated with the situational demands of the application process. Specifically, job applicants may feel under pressure to make a good impression on their prospective employer and they may subsequently present themselves as overly virtuous or defensive.

4.14.5 SUMMARY

Human fascination with the concept of personality lead to the nineteenth century invention of the first objective personality test.

As with those very early personality measures, many of today's objective personality inventories are self-reports. The clinician's first concern when utilizing an objective personality measure is whether or not a client can accurately reveal information about his or her personality through a self-report instrument. Many objective personality instruments incorporate indices of test-taking attitudes (e.g., the Lie scale of the MMPI/MMPI-2) which allow the clinician to gauge a client's level of insight and willingness to self-disclose. Additionally, research shows that clients who do cooperate with testing produce personality profiles that match external criteria (e.g., clinician's notes and observations regarding the patient). Essentially, most clients are able to reveal their personalities competently through self-report measures.

Once a client is able to self-disclose information regarding his or her personality, the scale scores that are produced appear to be quite stable over time. Thus, most of the objective personality measures manage to capture trait (as opposed to state) characteristics. Five factors exhibit influence on the stability of personality: (i) instrument characteristics, (ii) length of retest interval, (iii) operationalization of personality as a stable construct for test construction, (iv) the extent to which a particular personality construct is associated with stability, and (v) person variables of the test-taker.

Several objective personality measures are designed to assess adults in clinical settings. The most well-known and widely used objective personality inventory in clinical settings is the MMPI-2, which provides a comprehensive survey of personality characteristics and clinical problems. The Basic Personality Inventory (BPI) and the Personality Assessment Inventory (PAI) are alternatives to the MMPI-2, but neither is as widely used as the MMPI-2. Several other objective measures have been developed for specialized or focused clinical use including the Millon Clinical Multiaxial Inventory (MCMI) designed for making personality diagnoses; the Beck Depression Inventory (BDI), which assesses depressive ideation; the State-Trait Anxiety Inventory (STAI) designed to assess both long-term and short-term anxiety features; and the Whitaker Index of Schizophrenic Thinking (WIST) designed to measure the type of thought impairment that differentiates between schizophrenic and "normal" thought processes.

While certain objective personality measures were developed for and have been used widely in clinical setting, several inventories have gained use in research and other normal range assessment settings. Many researchers utilize inventories that measure the Big Five (or Five Factor Model, FFM) personality traits: Neuroticism (N), Extroversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C). The NEO Personality Inventory (NEO-PI) assesses N, E, and O and has been used as an index of the Big Five in research. The Multidimensional Personality Questionnaire (MPQ) assesses the broad personality domains of Positive Affectivity, Negative Affectivity, and Constraint. The MPQ contains two validity indexes (VRIN and TRIN) common to the MMPI-2, which make the MPQ appealing to many researchers. Other objective personality instruments that have been used widely in research include the Personality Research Form (PRF), a measure of several psychosocial needs; the 16PF, which measures global factors similar to the Big Five; and the California Psychological Inventory (CPI), designed to assess "everyday" concepts about personality.

Finally, professionals working in the fields of educational/vocational assessment and personnel selection have found use for several objective personality measures. The 16PF, the CPI, and, most commonly, the MMPI/MMPI-2 have been used in both of these settings. Objective personality instruments give professionals information about an individual's personality which would not be obtained through standard applications or interviews. Thus, professionals can use objective personality inventories as an efficient method of obtaining more information to aid them in their task of advising clients about educational/vocational decisions or advising employers in the selection of personnel.

Whether one needs a tool to assess adult personality in clinical, research, or industry settings, there is probably an objective personality inventory to fit the bill. Many of today's objective inventories offer the efficiency of providing a comprehensive assessment of personality functioning, and some even provide computerized interpretation of the personality profile. Perhaps most importantly, most of the objective personality inventories available commercially are standardized instruments that can assess adult personality validly and reliably, which means these instruments can be used over the course of a client's treatment to document goals for change—a feature that is becoming increasingly important to clinicians in this era of managed healthcare.

4.14.6 REFERENCES

Allard, C., & Mishara, B. L. (1995). Individual differences in stimulus intensity modulation and its relationship to two styles of depression in older adults. *Psychology and Aging, 10,* 395–403.

Alterman, A. I., Zaballero, A. R., Lin, M. M., Siddiqui, N., Brown, L. S., Jr., Rutherford, M. J., & McDermott, P. A. (1995). Personality Assessment Inventory scores of lower-socioeconomic African American and Latino methadone maintenance patients. *Assessment, 2,* 91–100.

Anderson, W. F. (1949). *Predicting success in nurses training.* Unpublished master's thesis, University of Nebraska, Lincoln, NE.

Appleby, T. L., & Haner, C. F. (1956). MMPI profiles of a college faculty group. *Proceedings of the Iowa Academy of Sciences, 53,* 605–609.

Applezweig, M. H. (1953). Educational levels and Minnesota Multiphasic profiles. *Journal of Clinical Psychology, 9,* 340–344.

Arbisi, P., & Ben-Porath, Y. S. (1995) An MMPI-2 infrequency scale for use with psychopathological populations: The Infrequency–Psychopathology Scale, F (p). *Psychological Assessment, 7,* 424–431.

Archer, R. P., Griffin, R., & Aiduk, R. (1995). Clinical correlates for ten common code types. *Journal of Personality Assessment, 65,* 391–408.

Assendorp, J. B. (1992). Beyond stability: Predicting inter-individual differences in intra-individual change. *European Journal of Personality, 6,* 103–117.

Bailey, D. M. (1988). Occupational therapy administrators and clinicians: Differences in demographics and values. *Occupational Therapy Journal of Research, 8,* 299–315.

Bailley, S. E., & Ross, S. R. (1996, May). *The effects of simulated faking on the five-factor (NEO-FFI).* Paper given at the 68th Annual Meeting of Midwestern Psychological Association, Chicago.

Barger, B., & Hall, E. (1964). Personality patterns and achievement in college. *Educational and Psychological Measurement, 24,* 339–346.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44,* 1–26.

Barthol, R. P., & Kirk, B. A. (1956). The selection of graduate students in public health education. *Journal of Applied Psychology, 40,* 159–163.

Bartram, D. (1992). The personality of UK managers: 16PF norms for short-listed applicants. *Journal of Occupational and Organizational Psychology, 65,* 159–172.

Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8,* 77–100.

Ben-Porath, Y. S., Butcher, J. N., & Graham, J. R. (1991). Contribution of the MMPI-2 scales to the differential diagnosis of schizophrenia and major depression. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3,* 634–640.

Ben-Porath, Y. S., McCully, E., & Almagor, M. (1993). Incremental validity of the MMPI-2 Content Scales in the assessment of personality and psychopathology by self-report. *Journal of Personality Assessment, 61,* 557–575.

Beutler, L. E., Nussbaum, P. D., & Meredith, K. E. (1988). Changing patterns of police officers. *Professional Psychology: Research and Practice, 19,* 503–507.

Beutler, L. E., Storm, A., Kirkish, P., Scogin, F., & Gaines, J. A. (1985). Parameters in the prediction of police officer performance. *Professional Psychology: Research and Practice, 16,* 324–335.

Blake, D. D., Penk, W. E., Mori, D. L., Kleespies, P. M., Walsh, S. S., & Keane, T. M. (1992). Validity and clinical scale comparisons between MMPI and MMPI-2 with psychiatric inpatients. *Psychological Reports, 70,* 323–332.

Boyle, G. J., & Lennon, T. J. (1994). Examination of the reliability and validity of the Personality Assessment Inventory. *Journal of Psychopathology and Behavioral Assessment, 16,* 173–187.

Burbeck, E., & Furnharm, A. (1985). Police officer selection: A critical review of the literature. *Journal of Police Science and Administration, 13,* 58–69.

Burch, J. W. (1995). Typicality range deficit in schizophrenics' recognition of emotion in faces. *Journal of Clinical Psychology, 51,* 140–150.

Burger, G. K., & Cross, D. T. (1979). Personality types as measured by the California Psychological Inventory. *Journal of Consulting and Clinical Psychology, 47,* 65–71.

Burgess, E. (1956). Personality factors of over- and under-achievers in engineering. *Journal of Educational Psychology, 47,* 89–99.

Bush, A. J., & Lucas, G. H. (1988). Personality profiles of marketing vs. R&D managers. *Psychology and Marketing, 5,* 17–32.

Butcher, J. N. (1989). *MMPI-2 scale correlates.* MMPI-2 Workshop Materials. Minneapolis MN: University of Minnesota Press.

Butcher, J. N. (1991). Screening for psychopathology: Industrial applications of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2). In J. Jones, B. D. Steffey, & D. Bray (Eds.), *Applying psychology in business: The manager's handbook.* Boston: Lexington.

Butcher, J. N. (1994). Psychological assessment of airline pilot applicants with the MMPI-2. *Journal of Personality Assessment, 62,* 31–44.

Butcher, J. N. (1995). Clinical use of computer-based personality test reports. In J. N. Butcher (Ed.), *Clinical personality assessment: Practical approaches* (pp. 78–94). New York: Oxford University Press.

Butcher, J. N., Berah, E., Ellertsen, B., Miach, P., Lim, J., Nezami, E., Pancheri, P., Derksen, J., & Almagor, M. (1998). Objective personality assessment: Computer-based MMPI-2 interpretation in international clinical settings. In C. Belar (Ed.), *Comprehensive clinical psychology: Sociocultural and individual differences.* New York: Elsevier.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring.* Minneapolis, MN: University of Minnesota Press.

Butcher, J. N., Graham, J. R., Williams, C. L., & Ben-Porath, Y. S. (1990). *Development and use of the MMPI-2 Content Scales.* Minneapolis, MN University of Minnesota Press.

Butcher, J. N. & Han, K. (1995). Development of an MMPI-2 scale to assess the presentation of self in a superlative manner: The S Scale. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment,* (Vol. 10, pp. 25–50). Hillsdale, NJ.: Erlbaum.

Butcher, J. N., Jeffrey, T., Cayton, T. G., Colligan, S., DeVore, J., & Minnegawa, R. (1990). A study of active duty military personnel with the MMPI-2. *Military Psychology, 2,* 47–61.

Butcher, J. N., & Rouse, S. V. (1996). Personality: Individual differences and clinical assessment. *Annual Review of Psychology, 47,* 87–111.

Butcher, J. N., & Williams, C. L. (1992). *MMPI-2 and MMPI-A: Essentials of clinical interpretation.* Minneapolis, MN: University of Minnesota Press.

Cattell, R. B. (1963). Personality, role, mood, and situation perception: A unifying theory of modulators. *Psychological Review, 70,* 1–18.

Cattell, R. B., Cattell, A. K. S., & Cattell, H. E. P. (1993). *Sixteen Personality Factor Questionnaire* (5th ed.). Champaign, IL: Institute for Personality and Ability Testing.

Centi, P. (1962). Personality factors related to college success. *Journal of Educational Research, 55,* 187–188.

Cerf, A. Z. (1947). Personality inventories. In J. P. Guilford (Ed.), *Printed classification tests.* Washington, DC: Army Air Force Aviation Psychology Program Research Reports.

Cerling, G. L. (1983). Selection of lay counselors for a church counseling center. *Journal of Psychology and Christianity, 2,* 67–72.

Chakrabarti, P. K., & Kundu, R. (1984). Personality profiles of management personnel. *Psychological Studies, 29,* 143–146.

Chick, D., Martin, S. K., Nevels, R., & Cotton, C. R. (1994). Relationship between personality disorders and clinical symptoms in psychiatric inpatients as measured by the Millon Clinical Multiaxial Inventory. *Psychological Reports, 74,* 331–336.

Chick, D., Sheaffer, C. I., Goggin, W. C., & Sison, G. F. (1993). The relationship between MCMI personality scales and clinician generated *DSM-III* personality disorder diagnoses. *Journal Personality Assessment, 61,* 264–276.

Clark, D. L. (1964). Exploring behavior in men's residence halls using the MMPI. *Personnel Guidance Journal, 43,* 249–251.

Clark, J. H. (1953). Grade achievement of female college students in relation to non-intellective factors: MMPI items. *Journal of Social Psychology, 37,* 275–281.

Clark, M. E. (1996). MMPI-2 negative treatment indicators content and content component scales: Clinical correlates and outcome prediction for men with chronic pain. *Psychological Assessment, 8,* 32–47.

Conley, J. J. (1985). Longitudinal stability of personality traits: A multi trait–multi method–multi occasion analysis. *Journal of Personality and Social Psychology, 49,* 1266–1282.

Cooper, C. L., & Green, M. D. (1976). Coping with occupational stress among Royal Air Force personnel on isolated island bases. *Psychological Reports, 39,* 731–734.

Costa, P. T., Jr., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment, 4,* 5–13.

Crookes, T. G., & Buckley, S. J. (1976). Lie score and insight. *Irish Journal of Psychology, 3,* 134–136.

Deniston, W. M., & Ramanaiah, N. V. (1993). California Psychological Inventory and the five-factor model of personality. *Psychological Reports, 73,* 491–496.

Dinning, W. D., & Evans, R. G. (1977). Discriminant and convergent validity of the SCL-90 in psychiatric inpatients. *Journal of Personality Assessment, 41,* 304–310.

Dunn, W. S., Mount, M. K., Barrick, M. R., & Ones, D. S. (1995). Relative importance of personality and general mental ability in managers' judgments of applicant qualifications. *Journal of Applied Psychology, 80,* 500–509.

Dyer, J. B., Sajwaj, T. E. G., & Ford, T. W. X. (1993, March). *MMPI-2 normative and comparative data for nuclear power plant personnel who were approved or denied security clearances for psychological reasons.* Paper presented at the 28th Annual Symposium on Recent Developments in the Use of the MMPI/MMPI-2, St. Petersburg, FL.

Egeland, B., Erickson, M., Butcher, J. N., & Ben-Porath, Y. S. (1991). MMPI-2 profiles of women at risk for child abuse. *Journal of Personality Assessment, 57,* 254–263.

Evans, R. G., & Dinning, W. D. (1980). A validation of Forms A and B of the Whitaker Index of Schizophrenic Thinking. *Journal of Personality Assessment, 44,* 416–419.

Fabricatore, J., Azen, S. P., Schoentgen, S., & Snibbe, H. (1978). Predicting performance of police officers using the Sixteen Personality Factor Questionnaire. *American Journal of Community Psychology, 6,* 63–70.

Fekken, G. C. (1985). The Whitaker Index of Schizophrenic Thinking. In D. J. Keyser & R. C. Sweetland (Eds.), *Test critiques* (Vol. III, pp. 717–725). Kansas City, KS: Test Corporation Of America.

Fekken, G. C., & Holden, R. R. (1991). The construct validity of person reliability. *Personality and Individual Differences, 12,* 69–77.

Ferris, G. R., Bergin, T. G., & Wayne, S. J. (1988). Personal characteristics, job performance, and absenteeism of public school teachers. *Journal of Applied Social Psychology, 18,* 552–563.

Finn, S. E. (1986). Stability of personality self-ratings over 30 years: Evidence for an age/cohort interaction. *Journal of Personality and Social Psychology, 50,* 813–818.

Fishkin, S. M., Lovallo, W. R., & Pishkin, V. (1977). Relationship between schizophrenic thinking and MMPI for process and reactive patients. *Journal of Clinical Psychology, 33,* 116–119.

Flynn, P. M. (1995). Issues in the assessment of personality disorder and substance abuse using the Millon Clinical Multiaxial Inventory (MCMI-II). *Journal of Clinical Psychology, 51,* 415–421.

Frick, J. W. (1955). Improving the prediction of academic achievement by use of the MMPI. *Journal of Applied Psychology, 39,* 49–52.

Fulkerson, S. C., Freud, S. L., & Raynor, G. H. (1958, February). The use of the MMPI in psychological evaluation of pilots. *Aviation Medicine* 122–128.

Fulkerson, S. C., & Sells, S. B. (1958). Adaptation of the MMPI for aeromedical practice norms for military pilots. *USAF School of Aviation Medicine Reports,* 58–128.

Furnham, A. (1991). Personality and occupational success: 16PF correlates of cabin crew performance. *Personality and Individual Differences, 12,* 87–90.

Garetz, F. K., & Tierney, R. W. (1962). Personality variables in army officer candidates. *Military Medicine, 127,* 669–672.

Geist, C. R., & Boyd, S. T. (1980). Personality characteristics of Army helicopter pilots. *Perceptual Motor Skills, 51*(1), 253–254.

Goorney, A. B. (1970). MMPI and MMPI scores, correlations and analysis for military aircrew population. *British Journal of Social and Clinical Psychology, 9,* 164–170.

Gottfredson, G. D., Jones, E. M., & Holland, J. L. (1993). Personality and vocational interests: The relation of Holland's six interest dimensions to five robust dimensions of personality. *Journal of Counseling Psychology, 40,* 518–524.

Gough, H. G. (1957). *Manual for the California Psychological Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Graham, J. R., & Ben-Porath, Y. S. (1990, June). *Congruence between the MMPI and MMPI-2.* Paper given at the 25th Annual Symposium on Recent Developments in the Use of the MMPI/MMPI-2, Minneapolis, MN.

Graham, J. R., & Butcher, J. N. (1988, March). *Differentiating schizophrenic and major affective disorders with the revised form of the MMPI.* Paper presented at the 23rd Annual Symposium on Recent Developments in the Use of the MMPI, St. Petersburg, FL.

Grigoriadis, S. (1993). *Sources of inconsistency on tests of psychopathology.* Unpublished doctoral dissertation, Queen's University, Kingston, ON.

Hargrave, G. E., & Hiatt, D. (1987, May). *Use of the MMPI to predict aggression in law enforcement officer applicants.* Paper presented at the 22nd Annual Symposium on Recent Developments in the Use of the MMPI, Seattle, WA.

Hargrave, G. E., & Hiatt, D. (1989). Use of the California Psychological Inventory in law enforcement officer selection. *Journal of Personality Assessment, 53,* 267–277.

Hartman, B. J. (1987). Psychological screening of law

enforcement candidates. *American Journal of Forensic Psychology, 5,* 5–10.
Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): 1. Construction of the schedule. *Journal of Psychology, 10,* 249–254.
Helson, R., & Moane, G. (1987). Personality change in women from college to midlife. *Journal of Personality and Social Psychology, 53,* 176–186.
Henney, A. S. (1975). Personality characteristics of a group of industrial managers. *Journal of Occupational Psychology, 48,* 65–67.
Hills, H. A. (1995). Diagnosing personality disorders: An examination of the MMPI-2 and MCMI-II. *Journal of Personality Assessment, 65,* 21–34.
Hindelang, M. J. (1972). The relationships of self-reported delinquency to scales of the CPI and MMPI. *Journal of Criminal Law, Criminology, and Police Science, 63,* 75–81.
Hjemboe, S., & Butcher, J. N. (1991). Couples in marital distress: A study of demographic and personality factors as measured by the MMPI-2. *Journal of Personality Assessment, 57,* 216–237.
Hoffman, R. G., & Davis, G. L. (1995). Prospective validity study: CPI work orientation and managerial potential scales. *Educational and Psychological Measurement, 55,* 881–890.
Holden, R. R., Fekken, G. C., Reddon, J. R., Helmes, E., & Jackson, D. N. (1988). Clinical reliabilities and validities of the Basic Personality Inventory. *Journal of Consulting and Clinical Psychology, 56,* 766–768.
Husband, S. D., & Iguchi, M. (1995). Comparison of MMPI-2 and MMPI clinical scales and high point scores among methadone maintenance clients. *Journal of Personality Assessment, 64,* 371–375.
Inch, R., & Crossley, M. (1993). Diagnostic utility of the MCMI-I and MCMI-II with psychiatric outpatients. *Journal of Clinical Psychology, 49,* 358–366.
Jackson, D. N. (1984). *Personality Research Form manual* (3rd ed.). Port Huron, MI: Research Psychologists Press.
Jackson, D. N. (1989). *Basic Personality Inventory manual.* Port Huron, MI: Research Psychologists Press.
Jackson, D. N., & Messick, S. (1971). *The Differential Personality Inventory.* London, ON: Authors.
Jaffe, L. T. & Archer, R. P. (1987). The prediction of drug use among college students from MMPI, MCMI, and sensation seeking scales. *Journal of Personality Assessment, 51,* 243–253.
Jennings, L. S. (1948). Minnesota Multiphasic Personality Inventory; differentiation of psychologically good and poor combat risks among flying personnel. *Journal of Aviation Medicine, 19,* 222.
Johnson, R. W., Flammer, D. P., & Nelson, J. G. (1975). Multiple correlations between personality factors and SVIB occupational scales. *Journal of Counseling Psychology, 22,* 217–223.
Keilen, W. G., & Bloom, L. J. (1986). Child custody evaluation practices: A survey of experienced professionals. *Professional Psychology: Research and Practice, 17,* 338–346.
Keller, L. S., & Butcher, J. N. (1991). *Use of the MMPI-2 with chronic pain patients.* Minneapolis, MN: University of Minnesota Press.
Khan, F. I., Welch, T., & Zillmer, E. (1998). MMPI-2 profiles of battered women in transition. *Journal of Personality Assessment, 60,* 100–111.
Kline, P., & Lapham, S. L. (1992). Personality and faculty in British universities. *Personality and Individual Differences, 13,* 855–857.
Knight, R. A., Epstein, B., & Zielony, R. D. (1980). The validity of the Whitaker Index of Schizophrenic Thinking. *Journal of Clinical Psychology, 36,* 632–639.
Koss, M. P., & Butcher, J. N. (1973). A comparison of psychiatric patients' self-report with other sources of clinical information. *Journal of Research in Personality, 7,* 225–236.
Krueger, R. F., Schmutte, P. S., Caspi, A., Moffitt, T. E., Campbell, K., & Silva, P. A. (1994). Personality traits are linked to crime among men and women: Evidence from a birth cohort. *Journal of Abnormal Psychology, 103,* 328–338.
Kuhne, A., Orr, S., & Baraga, E. (1993). Psychometric evaluation of post-traumatic stress disorder: The Multidimensional Personality Questionnaire as an adjunct to the MMPI. *Journal of Clinical Psychology, 49,* 218–225.
Lardent, C. L. (1991). Pilots who crash: Personality constructs underlying accident prone behavior of fighter pilots. *Multivariate Experimental Clinical Research, 10,* 1–25.
Laufer, W. S., Skoog, D. K., & Day, J. M. (1982). Personality and criminality: A review of the California Psychological Inventory. *Journal of Clinical Psychology, 38,* 562–573.
Lavin, P. F., Chardos, S. P., Ford, W. T., & McGee, R. K. (1987). The MMPI profiles of troubled employees in relation to nuclear power plant personnel norms. *Transactions of the American Nuclear Society, 54,* 146–147.
Lawrence, R. A. (1984). Police stress and personality factors: A conceptual model. *Journal of Criminal Justice, 12,* 247–263.
Leslie, B. A., Landmark, J., & Whitaker, L. C. (1984). The Whitaker Index of Schizophrenic Thought (WIST) and thirteen systems for diagnosing schizophrenia. *Journal of Clinical Psychology, 40,* 636–648.
Lorr, M. & Strack, S. (1994). Personality profiles of police candidates. *Journal of Clinical Psychology, 50,* 200–207.
Lubin, B., Larsen, R. M., & Matarazzo, J. D. (1984). Patterns of psychological test usage in the United States: 1935–1982. *American Psychologist, 39,* 451–454.
Lumsden, J. (1977). Person reliability. *Applied Psychological Measurement, 1,* 477–482.
MacCabe, S. P. (1987). Millon Clinical Multiaxial Inventory. In D. J. Keyser & R. C. Sweetland (Eds.), *Test Critiques Compendium* (pp. 304–315). Kansas City, KS: Test Corporation Of America.
MacLennan, R. N. (1991). *Personality Research Form annotated research bibliography.* Regina, AL: University of Regina, Department of Psychology.
McCrae, R. R., & Costa, P. T., Jr. (1990). *Personality in adulthood.* New York: Guilford.
Megargee, E. I. (1972). *The California Psychological Inventory handbook.* San Francisco: Jossey-Bass.
Millon, T. (1977). *Manual for the Millon Clinical Multiaxial Inventory.* Minneapolis, MN: National Computer Systems.
Millon, T. (1987). *Manual for the Millon Clinical Multiaxial Inventory-II.* Minneapolis, MN: National Computer Systems.
Millon, T. (1994). *Manual for the Millon Clinical Multiaxial Inventory-III.* Minneapolis, MN: National Computer Systems.
Millon, T., & Davis, R. (1995). Putting Humpty Dumpty together again: Using the MCMI in psychological assessment. In L. E. Beutler & M. R. Berren (Eds.), *Integrative assessment of adult personality* (pp. 240–279). New York: Guilford.
Morey, L. (1991). *Personality Assessment Inventory: Professional Manual.* Odessa, FL: Psychological Assessment Resources.
Murray, H. A. (1938). *Explorations in personality.* Cambridge, MA: Harvard University Press.
Patrick, J. (1993). Validation of the MCMI-I Borderline Personality Disorder scale with a well-defined criterion sample. *Journal of Clinical Psychology, 49,* 28–32.
Paunonen, S. V., Jackson, D. N., Trzebinski, J., & Fosterling, F. (1992). Personality structures across cultures: A multi method evaluation. *Journal of Personality and Social Psychology, 62,* 447–456.
Payne, F. D., & Wiggins, J. S. (1972). MMPI profile types and the self-report of psychiatric patients. *Journal of Abnormal Psychology, 79,* 1–8.
Payne, R. W. (1978). Review of Whitaker Index of Schizophrenic Thought. In O. K. Buros (Ed.), *The eighth mental measurements yearbook* (pp. 1146–1147). Highland Park, NJ: Gryphon Press.
Piedmont, R. L. (1993). A longitudinal analysis of burnout in the health care setting: The role of personal dispositions. *Journal of Personality Assessment, 61,* 457–473.
Pishkin, V., Lovallo, W. R., & Bourne, L. E. (1986). Thought disorder and schizophrenia: Isolating and timing a mental event. *Journal of Clinical Psychology, 42,* 417–424.
Reynolds, C. R. (1992). Review of the Millon Clinical Multiaxial Inventory-II. In J. J. Kramer & J. C. Conoley (Eds.), *The eleventh mental measurements yearbook* (pp. 533–535). Lincoln, NE: Buros Institute of Mental Measurements.
Saxe, S. J., & Reiser, M. (1976). A comparison of three police applicant groups using the MMPI. *Journal of Police Science and Administration, 4,* 419–425.
Schinka, J. A. (1995). Personality Assessment Inventory scale characteristics and factor structure in the assessment of alcohol dependency. *Journal of Personality Assessment, 64,* 101–111.
Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology, 78,* 966–974.
Schretlen, D. (1988). The use of psychological tests to identify malingered symptoms of mental disorder. *Clinical Psychology Review, 8,* 451–476.
Schuerger, J. M., Zarrella, K. L., & Hotz, A. S. (1989). Factors that influence the temporal stability of personality by questionnaire. *Journal of Personality and Social Psychology, 56,* 777–783.
Scogin, F., & Beutler, L. (1986). Psychological screening of law enforcement candidates. In P. A. Keller & L. G. Ritt (Eds.), *Innovations in clinical practice* (Vol. 5, pp. 317–330). Sarasota, FL: Professional Resources Exchange.
Scogin, F., & Reiser, M. (1976). A comparison of three police applicant groups using the MMPI. *Journal of Police Science and Administration, 4,* 419–425.
Soldz, S., Budman, S., Demby, A., & Merry, J. (1993). Diagnostic agreement between the Personality Disorder Examination and the MCMI-II. *Journal of Personality Assessment, 60,* 486–499.
Spielberger, C. D. (1977). Anxiety, theory and research. In B. Wolman (Ed.) *International encyclopedia of neurology, psychiatry, psychoanalysis, and psychology.* New York: Human Sciences Press.
Spielberger, C. D. (1976). Stress and anxiety and cardiovascular disease. *Journal of South Carolina Medical Association* (Supplement 15), *72,* 15–22.
Spielberger, C. D. (1983). *Manual for the State Trait Anxiety Inventory: STAI (form Y).* Palo Alto, CA: Consulting Psychologists Press.
Spielberger, C. D., Gorsuch, R. L., & Lushene, R. D. (1970). The STAI: Manual for the State–Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press.
Spielberger, C. D., Ritterband, L. M., Sydeman, S. J., Reheiser, E. C., and Unger, K. K. (1995). Assessment of emotional states and personality traits: Measuring psychological vital signs. In J. N. Butcher (Ed.), *Clinical personality assessment. Practical approaches* (pp. 43–58). New York: Oxford University Press.
Tellegen, A. *Brief manual for the Differential Personality Questionnaire.* Unpublished manuscript.
Topp, B. W., & Kardash, C. A. (1986). Personality, achievement, and attrition: Validation in a multiple-jurisdiction police academy. *Journal of Police Science and Administration, 14,* 234–241.
Walsh, W. B. (1974). Consistent occupational preferences and personality. *Journal of Vocational Behavior, 4,* 145–153.
Watkins, C. E. (1996). On Hunsley, Harangue, and Hoopla. *Professional Psychology: Research and Practice, 27,* 316–318.
Weed, N. C., Butcher, J. N., Ben-Porath, Y. S., & McKenna, T. (1992). New measures for assessing alcohol and drug abuse with the MMPI-2: The APS and AAS. *Journal of Personality Assessment, 58,* 389–404.
Whitaker, L. C. (1973). *The Whitaker Index of Schizophrenic Thinking.* Los Angeles: Western Psychological Services.
Whitaker, L. C. (1980). *Objective measurement of schizophrenic thinking: A practical and theoretical guide to the Whitaker Index of Schizophrenic Thinking.* Los Angeles: Western Psychological Services.
Whitbourne, S. K., Zuschlag, L. B., Elliot, L. B., & Waterman, A. S. (1992). Psychosocial development in adulthood: A 22-year sequential study. *Journal of Personality and Social Psychology, 63,* 260–271.
Wiggins, J. S. (1969). Content dimensions in the MMPI. In J. N. Butcher (Ed.), *MMPI: Research developments and clinical applications* (pp. 127–180).
Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment.* Reading, MA: Addison-Wesley.
Williams, J. E., Munick, M. L., Saiz, J. L., & Formy-Duval, D. L. (1995). Psychological importance of the "Big Five": Impression formation and context effects. *Personality and Social Psychology Bulletin, 21,* 818–826.
Windle, C. (1954). Test–retest effect on personality questionnaires. *Educational and Psychological Measurement, 14,* 617–633.

Seaton, B., & Allen, J. (1996, March). *Interscorer reliability of Rorschach structural summary data.* Poster session presented at the annual meeting of the Society for Personality Assessment, Denver, CO.

Smith, D., & Dumont, F. (1995). A cautionary study: Unwarranted interpretations of the Draw-A-Person test. *Professional Psychology, 3,* 298–303.

Stinnett, T. A., Havey, J. M., & Oehler-Stinnett, J. (1994). Current test usage by practicing school psychologists: A national survey. *Journal of Psychoeducational Assessment, 12,* 331–350.

Stone, H. K., & Dellis, N. P. (1960). An exploratory investigation into the levels hypothesis. *Journal of Projective Techniques, 24,* 333–340.

Sutton, P. M., & Swenson, C. H. (1983). The reliability and concurrent validity of alternative methods for assessing ego development. *Journal of Personality Development, 47,* 468–475.

Swensen, C. H. (1957). Empirical evaluations of human figure drawings. *Psychological Bulletin, 54,* 431–466.

Swensen, C. H. (1968). Empirical evaluations of human figure drawings: 1957–1966. *Psychological Bulletin, 70,* 20–44.

Tharinger, D. J., & Stark, K. (1990). A qualitative versus quantitative approach to the Draw-a-Person and Kinetic Family Drawing: A study of mood- and anxiety-disorder children. *Psychological Assessment, 2,* 365–375.

Thomas, A. D., & Dudek, S. Z. (1985). Interpersonal affect in TAT responses: A scoring system. *Journal of Personality Assessment, 49,* 30–37.

Vane, J. R. (1981). The Thematic Apperception Test: A review. *School Psychology Review, 1,* 319–336.

Watkins, C. E., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology, 26,* 54–60.

Weiner, I. B. (1977). Projective tests in differential diagnosis. In B. B. Wolman (Ed.), *International encyclopedia of neurology, psychiatry, psychoanalysis, and psychology* (pp. 112–116). Princeton, NJ: Van Nostrand Reinhold.

Weiner, I. B. (1986). Assessing children and adolescents with the Rorschach. In H. M. Knoff (Ed.), *The assessment of child and adolescent personality* (pp. 141–171). New York: Guilford Press.

Weiner, I. B. (1993). Clinical considerations in the conjoint use of the Rorschach and the MMPI. *Journal of Personality Assessment, 60,* 148–152.

Weiner, I. B. (1994). Rorschach assessment. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome evaluation* (pp. 249–278). Hillsdale, NJ: Erlbaum.

Weiner, I. B. (1995a). Methodological considerations in Rorschach research. *Psychological Assessment, 7,* 330–337.

Weiner, I. B. (1995b). Psychometric issues in forensic applications of the MMPI-2. In Y. S. Ben-Porath, J. R. Graham, G. C. N. Hall, R. D. Hirschman, & M. S. Zaragoza (Eds.), *Forensic applications of the MMPI-2* (pp. 48–81). Thousand Oaks, CA: Sage.

Weiner, I. B. (1996). Some observations on the validity of the Rorschach Inkblot Method. *Psychological Assessment, 8,* 206–213.

Weiner, I. B. (1997). Current status of the Rorschach Inkblot Method. *Journal of Personality Assessment, 68,* 5–19.

Weiner, I. B., & Exner, J. E., Jr. (1991). Rorschach changes in long-term and short-term psychotherapy. *Journal of Personality Assessment, 56,* 453–465.

Westen, D., Klepser, J., Ruffins, S. A., Silverman, M., Lifton, N., & Boekamp, J. (1991). Object relations in childhood and adolescence: The development of working representations. *Journal of Consulting and Clinical Psychology, 59,* 400–409.

Westen, D., Lohr, N., Silk, K., Kerber, K., & Goodrich, S. (1985). *Object relations and social cognition TAT scoring manual.* Ann Arbor, MI: University of Michigan.

Westen, D., Ludolph, P., Block, M. J., Wixom, J., & Wiss, F. C. (1990). Developmental history and object relations in psychiatrically disturbed adolescent girls. *American Journal of Psychiatry, 147,* 1061–1068.

Westen, D., Ludolph, P., Lerner, H., Ruffins, S., & Wiss, F. C. (1990). Object relations in borderline adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 29,* 338–348.

Worchel, F. F., Rae, W. A., Olson, T. K., & Crowley, S. L. (1992). Selective responsiveness of chronically ill children to assessments of depression. *Journal of Personality Assessment, 59,* 605–615.

4.16 Assessment of Schema and Problem-solving Strategies with Projective Techniques

HEDWIG TEGLASI
University of Maryland, College Park, MD, USA

4.16.1 INTRODUCTION 460

4.16.2 PATTERNS OF USE 460

4.16.3 BASIC ASSUMPTIONS OF PROJECTIVE TECHNIQUES AND RESEARCH FROM OTHER PSYCHOLOGY SUBFIELDS 461
- *4.16.3.1 Schema Theory and Projective Techniques* 462
- *4.16.3.2 Unconscious Processing of Information* 464
- *4.16.3.3 Inert Knowledge vs. Usable Knowledge* 465
- *4.16.3.4 Affective and Motivational Influences on Cognition* 465
- *4.16.3.5 Narrative Psychology and Thematic Apperception Techniques* 466

4.16.4 LEVELS OF PERSONALITY ASSESSED BY PROJECTIVE AND QUESTIONNAIRE METHODS 467

4.16.5 SELF-REPORT VS. PROJECTIVE PERSONALITY TESTING 470

4.16.6 CONTRIBUTION OF PROJECTIVE TECHNIQUES TO DIAGNOSIS AND INTERVENTION 470

4.16.7 PROJECTIVE TECHNIQUES AS PERFORMANCE MEASURES OF PERSONALITY 472
- *4.16.7.1 Total Personality* 474
- *4.16.7.2 Many Correct Solutions* 476
- *4.16.7.3 Generalize to Different Criteria* 476
- *4.16.7.4 Differences in Conditions of Learning and Performance* 477

4.16.8 SPECIFIC PROJECTIVE TECHNIQUES 477
- *4.16.8.1 Stimulus* 477
- *4.16.8.2 Response* 477
- *4.16.8.3 Task Demand* 478
- *4.16.8.4 Interpretation* 478

4.16.9 THEMATIC APPERCEPTIVE TECHNIQUES 479
- *4.16.9.1 Response* 479
- *4.16.9.2 Stimuli* 479
- *4.16.9.3 Interpretation* 480

4.16.10 RORSCHACH TECHNIQUE 481
- *4.16.10.1 Stimulus Features and Response Parameters* 481
- *4.16.10.2 Interpretation* 483

4.16.11 DRAWING TECHNIQUES 482
- *4.16.11.1 Interpretation* 484

4.16.12 CASE ILLUSTRATION 484

4.16.13 VALIDATION ISSUES 487
4.16.13.1 Reliability 488
4.16.13.1.1 Scorer reliability 488
4.16.13.1.2 Decisional reliability 488
4.16.13.1.3 Test–retest reliability 488
4.16.13.1.4 Internal consistency 488
4.16.13.2 Construct Validation 489
4.16.13.3 Multitrait, Multimethod Validation 489
4.16.13.4 Criterion Validation 489
4.16.13.5 Part–Whole Relationships 490
4.16.13.6 Convergence of Psychometric and Conceptual Treatment of Data 490
4.16.13.6.1 Normative 491
4.16.13.6.2 Nomothetic 491
4.16.13.6.3 Case study 491
4.16.14 FUTURE DIRECTIONS 492
4.16.15 SUMMARY 493
4.16.16 REFERENCES 495

4.16.1 INTRODUCTION

Clinical assessment practices typically lag behind the conceptualizations psychologists use to understand adaptive and maladaptive patterns in human functioning. We recognize that personality is a functional whole, but cling to assessment procedures that resemble the proverbial blind men and the elephant. Each method of assessment pertains to one dimension or level of personality, yet we expect different sources of data to confirm one another rather than to reveal the different patterns contributing to the larger mosaic. A clarification of the part–whole relationships with regard to personality assessment promotes an understanding of the place of projective techniques in a comprehensive evaluation. The aims of this chapter are: (i) to facilitate the process of integration of projective techniques with self-report approaches to personality assessment; (ii) to reframe projective techniques as performance measures of personality that differ in their demands from other tasks within the battery; (iii) to show that the central tenets of projective techniques embodied in the projective hypothesis are supported by empirical research in various psychology subfields; and (iv) to demonstrate the utility of projective techniques for assessing schema (the internal representation of reality) and problem-solving strategies to meet the reality demands posed by the task.

That projective techniques do not belong to a particular theory of personality was asserted early on (Auld, 1954). Projective methods yield open-ended responses, and their interpretation can accommodate changing paradigms in various subdisciplines, such as the cognitive and neurosciences as well as in the study of emotion and social cognition. It is easy to say that the framework to understand and interpret responses to projective tasks should coincide with key concepts in the various subdisciplines, but the task of continuously updating the conceptual frameworks that are applied to projective methods remains an ongoing challenge.

Responses to projective tests yield information about the internal structure of personality and provide the context for understanding overt behavior and symptomatology. This emphasis on inner structure is currently shared by social, cognitive, and clinical psychology. Newer appreciation is emerging of the wide role of emotions and unconscious processes (e.g., Bargh, 1994; Clore, Schwarz, & Conway, 1994), as well as of meaning structures (e.g., Bruner, 1986) underlying behavioral expression. Projective methods reveal the structure and process of personality that are largely unconscious, and they differ in important ways from checklists or self-report interviews. In addition, the competencies required by projective tasks can be distinguished from those demanded by other measures in the typical assessment battery such as the Wechsler scales.

Therefore, projective techniques can be viewed as performance measures of personality that reveal inner structures or schema as resources needed in daily life. Each projective method sets a task that requires the application of these structures to respond adaptively. All factors contributing to personality development are involved in shaping these inner structures of meaning. Therefore, the interpretation of responses to projective techniques promotes the integration of knowledge from various subdisciplines.

4.16.2 PATTERNS OF USE

Despite criticisms of projective tests, there continues to be broad interest in the Rorschach and the Thematic Apperception Test (TAT; Butcher & Rouse, 1996). Most clinical psychology doctoral training programs include formal instruction in the Rorschach (Piotrowski & Zalewski, 1993). In addition, the majority of internship sites approved by the American Psychological Association (APA) value knowledge in Rorschach technique (Durand, Blanchard, & Mindell 1988). The Exner Comprehensive System (Exner, 1993; Exner & Weiner, 1995) has become widely accepted because it provides a more reliable and objective basis for interpretation than was previously available. Surveys find the Rorschach, TAT, and various drawing methods among the top 10 most frequently used assessment techniques (Archer, Marnish, Imhof, & Piotrowski, 1991; Piotrowski & Keller, 1989; Piotrowski, Sherry, & Keller, 1985; Watkins, Campbell, & McGregor, 1988; Watkins, Campbell, Nieberding, & Hallmark, 1995). Interest in the Rorschach and TAT is also evident outside of the USA (Weiner, 1994).

The contrast between the negative evaluation of projective methods by researchers and their continued popularity with practitioners (Piotrowski, 1984) begs for a scrutiny of the assumptions and methods applied to judging the utility of these techniques. This state of affairs calls for an analysis not only of psychometric properties but also of the conceptual and methodological underpinnings pertaining to projective testing. Experienced practitioners continue to use these techniques because they find them helpful. For instance, an advantage of the Rorschach is that the validity of its administration is independent of the client's reading level (Archer et al., 1991). Shneidman's (1951) edited volume presented the conclusions of 16 clinicians who blindly interpreted the same TAT protocol using their own preferred method. Despite different approaches and foci, they were remarkably accurate and consistent with each other. Also included in this volume were Klopfer's conclusions based on a blind interpretation of the same individual's Rorschach record which was highly consistent with those based on the TAT protocols. It should be noted that the interpreters all had one thing in common—they were "experts." In light of what we know about how experts perform (e.g., Chi, Glaser, & Farr, 1987), we would not expect to replicate such findings with more recently trained individuals.

In evaluating the evidence of reliability and validity in the clinical use of the TAT, the training and experience of interpreters was a primary consideration (Karon, 1981). Research designs that isolate response variables and subject them to separate statistical analyses do not mirror the interpretation of projective methods in clinical practice. The clinician simultaneously evaluates multiple dimensions of the response in reference to a psychological construct rather than adding individual response variables. More sophisticated validation efforts are needed that focus-on the integrative approach rather than piecemeal summation of isolated response characteristics (e.g., Handler & Habenicht, 1994). Finally, more refined approaches to validation suggest that different facets of psychological constructs are measured by different techniques. For instance, a comprehensive meta-analysis indicated that projective measures of achievement motivation predict long term behavioral trends, whereas self-report measures predict immediate choices (Spangler, 1992). Refinements in conceptualization and approaches to validation are calling into question earlier criticism based on psychometric shortcomings.

4.16.3 BASIC ASSUMPTIONS OF PROJECTIVE TECHNIQUES AND RESEARCH FROM OTHER PSYCHOLOGY SUBFIELDS

The term "projective technique" was coined by Frank (1939) to describe tests using ambiguous stimuli and/or tasks that are less obvious in their intent and are therefore less subject to faking. The aim of any projective technique is to set a task that allows the individual to express characteristic ways of perceiving and organizing experiences. The uniqueness of the individual's style of responding is particularly evident when stimuli are ambiguous and there is no ready response. Under such conditions, there are many possible ways to approach the task or situation, and the person must actively organize the response. A central assumption of all projective methods is that stimuli from the environment are perceived and organized by the individual's specific needs, motives, feelings, perceptual sets, and cognitive structures, and that in large part this process occurs automatically and outside of awareness (Frank, 1948).

These assumptions about projective testing, first elaborated by Frank (1939) in his analysis of the projective hypothesis, are compatible with converging trends across various subfields in psychology pointing to the role of previously organized inner structures or mental sets in the interpretation of stimuli. Since the cognitive revolution of the 1960s, psychology has undergone a significant transformation by moving away from an emphasis on stimulus–response units to describe human perception and action to increasing emphasis on how humans impose

meaning and organization on their experiences (Singer & Salovey, 1991). Within cognitive psychology, simple associational models (e.g., paired associate learning) have become less important in describing ongoing information processing than the organized meaning structures, such as schema or scripts that are brought to bear on the interpretation of experiences (Piaget, 1954; Taylor & Crocker, 1981). At the same time, developments in social psychology (Abelson, 1976; Festinger, 1957; Heider, 1958; Kelley, 1967) led to the exploration of organized meaning structures as the fundamental bases of social action (Wyer & Srull, 1994). In the psychodynamic paradigm, the emergence of object relations theory (Blatt & Wild, 1976) implicitly relies on internal structures similar to schema to guide information processing about the self and other. The considerable work in cognitive, social, personality, and clinical psychology bearing on schema theory and supporting the projective hypothesis is briefly highlighted next.

4.16.3.1 Schema Theory and Projective Techniques

Earlier cognitive work within clinical psychology focused on the role of self-statements and inner thoughts. But clinical scientists interested in the role of cognition have increasingly adopted an information-processing approach to clinical phenomena (Ingram, 1986). This approach focuses on the schema that provide rules guiding behavior in social relationships and appear to influence how information about relationships is stored in memory (Fiske, Haslam, & Fiske, 1991). Well-learned, organized knowledge structures (schema or scripts) have considerable impact on memory storage and retrieval. Projective tasks require the superimposition of previously acquired schematic structures to the perception of stimuli and the organization of the response. Acklin (1994) presented a reformulation of the responses to the Rorschach on the basis of schema theory and information processing. Within this framework, emphasis is placed on the nature and adaptiveness of the schema activated by the stimulus properties.

Memory structures such as schema or scripts that store information about situations, persons, and events guide the interpretation of experiences by providing criteria for regulating attention to lend focus to the process of encoding, storage, and retrieval of information in specific domains (Taylor & Crocker, 1981). Event schema refer to the type of script that organizes understanding of a sequence of events in a stereotyped or routine situation such as ordering dinner in a restaurant (Abelson, 1981). Such a "script is a set of expectations about what will happen next in a well understood situation" (Schank, 1990, p. 7). As long as we know what script others are following, we know how to act and to predict the actions of others. Upon entering a restaurant, we are seated, order from the menu, receive our meal, pay, and depart. This process of comparing experience with the internal schema enables us to perceive rapidly unstated information and to anticipate what will happen next. Some individuals may view situations for which they have no scripts (novel) as challenging, whereas others are stymied (e.g., "speechless") in such conditions.

Person schema include information about the self, others, and their interactions. They deal with the individual's rules for predicting, interpreting, responding to, and controlling affectively charged encounters (Tomkins, 1979). For Tomkins (1979, 1987, 1991), emotion is a key organizer of personal scripts. Once the script has been formed, the script organizes and modifies new experiences to fit the preexisting structure. Person schema develop through the individual's synthesis of past experiences according to the individual's style of processing information (Horowitz, 1991). Therefore the construct of person schema brings together models of perception, cognition, memory, affect, action, and feedback. Furthermore, it has the potential to integrate psychodynamic formulations and cognitive perspectives in understanding personality (Horowitz, 1991; Stein & Young, 1992). This process of interpreting new information according to previously acquired sets takes place without conscious awareness and may also lead to systematic distortions in perception, interpretation, and action in interpersonal encounters.

Causal schema give rise to attributions about the causes of events. Schema theory and attribution theory overlap in that attribution theory maintains that individuals are motivated to make causal inferences about experiences and that attributions are made in ways that are congruent with existing schema of self and other, and of the assumed relationships between causes and effects (Tunis, 1991). The relationship between such causal attributions and depression has been studied.

Evidence for the existence of schema is seen in their consequences for information processing, particularly in how information is remembered and retrieved. Schema-driven information processing involves the application of organized knowledge structures akin to theories that govern the perception and interpretation of facts. Thus, schema function to allow general

knowledge to influence perception of specific experiences. Schema increase efficiency in identifying perceptions, organizing them into manageable units, filling in missing information, and selecting a strategy for how to obtain further information if needed. Because schema guide behavior in adaptive and maladaptive ways, their study pertains to clinical psychology as well as to the general study of personality, cognition, and social perception. Understanding the recurrence of maladaptive interpersonal patterns is facilitated by studying the operation of schema along with the realistic properties of the current situation (Horowitz, 1991).

The overlap between schema or script theory and the projective hypothesis is evident in their common features. The development of schema or scripts and their retrieval from memory represent an unconscious process through which past perceptions influence the interpretation of current situations. These are precisely the processes that are assessed with projective techniques. The focus of projective methods is on the application of schema to the task demands (e.g., how schema influence the response process) and, by extension, to life challenges requiring similar adaptations.

Even though schema are structures of the mind, they must be sufficiently malleable to adapt to new situations and new configurations of events (Rummelhart, Smolensbus, McClelland, & Hinton, 1986). Therefore, it is appropriate to view schema as emerging when they are needed "from the interaction of large numbers of much simpler elements all working in concert with one another" (p. 20). A distinction has been made between enduring schema and working models (Horowitz, 1991). The former are intrapsychic meaning structures containing generalized formats of knowledge that can be activated by other mental activities related to that knowledge such as motivational concerns (Fiske & Taylor, 1984). Working models combine internal and external sources of information, such as when the individual is actively contemplating an interpersonal situation or task such as the TAT. Working models actively integrate stimuli from a current situation with past knowledge through triggering of an associative network of ideas drawn from enduring person schema.

A person's repertoire may include several different enduring person schema in relation to a given type of relationship, situation, or activity. The working model may incorporate different elements from each of the person schema. The wider the repertoire of enduring person schema, the more flexibility in constructing working models of interpersonal situations. Discrepancies between working models and active, enduring schema may contribute to upsetting emotional experiences. Working models may not match the actual qualities of current social situations, leading to errors in judgment and behavior and to subsequent negative emotions. The aim of many psychotherapeutic techniques is to promote conscious awareness of unconscious, schematic functions. Such awareness may permit the individual to actively override some of the influence of the unconscious information processing.

Scripts or schema may be further organized into metascripts (Singer & Salovey, 1991) that reflect the individual's style of dealing with scripts. These superordinate schema may explain resilience because they influence how individuals reshape their schema when confronted with daily stress, unforeseen failure, or unexpected upheaval. Those with more complex schema may be more willing to replay negative information to explore a variety of alternative schema or scripts. More flexible schema involve the realization that current ideas that seem so clear and self-evident at the moment may change as a function of new experience. This process cannot be taught directly, but is a byproduct of experiencing changes in perception, understanding, and feeling in light of new information. Consequently, it is important to assess the style of synthesizing experience through the organization and reorganization of schema.

Analysis of responses elicited by projective techniques can be informed by models of schema development. Abelson (1976) assumes that schema begin as representations of single concrete examples and become more abstract. Rudimentary schema are compilations of single, concrete examples that are used to make snap judgments about seemingly similar instances. Stereotypic schema encompass the most representative features of events, persons, or groups. More abstract schema recognize inconsistencies between reality and the activated schema. The individual is aware of complexities and ambiguities rather than simply encoding schema–congruent information. Highly developed schema permit the individual to engage in complex information processing seemingly effortlessly and without awareness. The more abstract the schema, the more flexibly they can be applied because they include conditional and inferential concepts, abstract rules, and affective information. The progression from rudimentary to more complex schema is the product of experience and the degree of active, strategic effort brought to the structuring and organization of experience. Schema theory recognizes the role of affect, motivation, and biological

factors (genetics, temperament), as well as environmental influences (stress and supports) as they interact with cognitive processes. The shaping of schema, especially social scripts by culture, is also acknowledged and has been studied by anthropologists and linguists (Quinn & Holland, 1987).

Responses to projective methods may be evaluated according to aspects of schema described above. When telling stories to picture stimuli, some individuals superimpose associative elements without an organizational network; others impose stereotypes that may or may not precisely fit the stimulus configurations; still others draw creatively on various elements of their experiences to tell a cohesive story that captures the gist of the stimulus and complies with the instructions. If an individual is presented with a TAT picture that cannot be readily explained by a stereotypic script or an intact schema, then the respondent is called upon to actively construct a script to fit the situation. Similarly, when faced with inkblots that are rough approximations of real objects, the individual must remain flexible in applying schema to answer the question "What might this be." Those who rely on more rudimentary schema will make associations directly from discrete portions of the stimuli. However, those who apply more cohesive schema find a meaningful framework with which to integrate disparate or complex stimuli.

Two aspects of long-term memory structures (schema), procedural and declarative knowledge (Anderson, 1983; Kihlstrom, 1984), are relevant to the interpretation of projective responses. Procedural knowledge refers to unconscious processes or skills such as the structure of language, the organization of music, or other implicit rules that order information or perception. Declarative schematic structures entail the recall of factual information such as names, locations, and historical events. Responses to projective techniques reveal the implicit organization of knowledge as well as the content that enters awareness in response to the stimuli presented. Interpretive approaches can focus on the organization of schema by examining the structural aspects of responses and the sequence of ideas expressed, as well as by analyzing the content.

4.16.3.2 Unconscious Processing of Information

The idea that human behavior is based on a considerable storehouse of organized knowledge structures operating outside of conscious awareness is now widely accepted by cognitive psychologists (Bargh, 1994). Since the 1970s, increasing evidence suggests that information not accessible to conscious awareness influences memory, perception, and thinking (Kihlstrom, 1990). Efforts to refine assumptions about unconscious processes continue across the various subfields of psychology and are not restricted to projective testing or even to clinical psychology.

Social cognitive psychology (social cognition) has become increasingly interested in unconscious thought processes of mental representations of situations, self, and others (schema) and in the interaction of cognition and affect (Wyer & Srull, 1994). Attitudes, expectations, or schema that are strong enough to be automatically activated have been described in terms of their "chronic accessibility" (Fazio, Sanbonmatsu, Powell, & Kardes, 1986). Chronically accessible schema may be cued by emotions arising in ambiguous situations, including reactions to projective stimuli. According to Westen (1993), "Research on chronic schema accessibility has confirmed the projective hypothesis that in ambiguous situations enduring interests, concerns, needs, and ways of experiencing reality are likely to be expressed" (p. 381). Bargh (1994) suggests that temporarily accessible schema due to a recent experience can mimic chronic accessibility and constitute a pitfall for studies of enduring cognitive structures. Indeed, this pitfall is common in the use of personality tests, be they questionnaire or projective devices. The distinction between state and trait variables acknowledges states as temporary conditions of mentality, mood, levels of arousal or drive, and these states may be related to temporarily evoked schema. One of the advantages of projective measures is the opportunity to examine both the structure and content of schema even if temporarily evoked. The content of the schema may be more susceptible to influence by such variations in experience than the structure. However, data are needed to support this contention.

A dramatic demonstration of the effect of a recent experience occurred one semester when the author's students administered assessment batteries to incarcerated youths (generally between the ages of 18 and 22) to determine educational strategies. A great number of TAT stories ended by the characters "talking about it and solving their problem." After some investigation, it was learned that a communication program had been recently instituted to teach the value of resolving conflict by talking things out. A close inspection of the stories themselves revealed that the schema about communication were not well developed but verbalized as a stock ending apart from other components of the problem-solving process. For example, the problems were poorly understood, and characters' feelings, perspectives, and intentions were vaguely or inaccurately deciphered, if at all. The idea for talking about the problem became accessible to conscious awareness through training as declarative knowledge but without the understandings or procedural knowledge needed for its implementation. Conscious awareness of the benefits of communication was not combined with the implicit and largely unconscious understandings and affective reactions that promote such interactions. This example illustrates the importance of studying the inner logic and cohesiveness of schematic structures to discover how story content derived from recent experiences (including vicarious ones such as books or movies) is integrated with other aspects of the task demand such as accurately perceiving the stimuli and following instructions. Both declarative and procedural knowledge must be addressed to overcome the pitfall of attributing too much weight to content that is not meaningfully incorporated into guiding structures.

4.16.3.3 Inert Knowledge vs. Usable Knowledge

Inert knowledge has been described as information that the person knows, but does not apply unless explicitly cued or prompted to do so (Bransford, Franks, Vye, & Sherwood, 1989). The reason why a person fails to use relevant knowledge to solve particular problems is that this information does not spontaneously come to awareness in the context of that problem, or that the schema for applying the information are not sufficiently developed. The manner in which knowledge is encoded and organized in the first place determines its subsequent accessibility and applicability in other situations.

Effective learning in general requires active strategies to organize and recall information such as rehearsal or breaking down the task into manageable units. Torgesen (1977), who introduced the concept of the inactive learner, reported that learning-disabled students exhibit a learning style characterized by minimal planning and limited use of strategies. Individuals may process information from one domain more thoroughly than in others, or the inactive stance may be more pervasive. This inactive style is manifested by processing information exactly as it comes in (memorization) without organizing and restructuring it. In the absence of active, effortful, strategic processing of information, the knowledge may be dormant in memory and not available for use unless externally prompted.

Different types of tasks are needed to differentiate knowledge that is used spontaneously to solve problems from knowledge that is available only when cued by the context. Everyday behavior requires spontaneous use of prior knowledge to make social judgments or other decisions; the possession of factual knowledge is far removed from problem solving in everyday encounters. The TAT stories of the incarcerated youths referred to earlier shows how knowledge may be stated without being embedded in a usable framework. Projective measures differ from typical cognitive tasks and self-report measures of personality in ways that have relevance for distinguishing between inert and useful knowledge. Rather than being asked to provide specific information about the self, the respondent is provided with a relatively unstructured task and asked to give an open-ended response that is evaluated by a professional. Likewise, solving verbally communicated problems such as those on intelligence tests does not resemble the conditions of everyday situations. Such problems tend to contain words that act as cues for accessing relevant knowledge, whereas everyday problem-solving requires the individual to size up relevant features of the situation to notice that there is a problem (Sternberg, 1985). Real problems are accompanied by emotions that shape the way the individual thinks about the problem. Individuals may verbalize several acceptable alternatives for handling a situation, if prompted. Nevertheless, when faced with real-life dilemmas, they may be unable to translate this knowledge into appropriate action. The projective tests reveal the spontaneous accessibility of the schema that guide the application of what comes to awareness.

4.16.3.4 Affective and Motivational Influences on Cognition

Most studies in this area focus on the way mood or emotion influences memory and judgment (Isen, 1984, 1987) and on unconscious processing of information (Uleman & Bargh, 1989). Epstein (1994) describes the impact of emotion on thinking and argues for the existence of two fundamentally different but interactive modes of information processing: a rational system and an experiential system driven by emotion. Thinking in the experiential system is shaped by emotion and tends to occur without deliberate effort or conscious awareness. The rational system is characterized by a more deliberate process of thinking and of

acquiring information such as through textbooks and direct teaching. Epstein explains that stories and anecdotes spice up an otherwise dry lecture because they increase emotional engagement and thereby appeal to the experiential system.

Emotions dramatically alter the thought process and influence behavior. Thinking is transformed by intense emotion in the direction of being categorical, personal, concrete, unreflective, and action oriented. The product of such thinking is likely to be considered as self-evident and not requiring proof (Epstein, 1994). Depressed affect is associated with preference for immediate gratification in place of delayed but more substantial rewards (Wertheim & Schwartz, 1983). Anxiety tends to bias the interpretation of ambiguous text (MacLeod & Cohen, 1993). Chronically experienced negative emotions appear to have a cumulative influence on the development of cognitive structures amenable to assessment with projective techniques. Children temperamentally prone to experience negative affect told stories to TAT stimuli that reflected less complex information processing and more reactive self-regulatory styles dominated by more immediate concerns than their less emotional peers (Bassan-Diamond, Teglasi, & Schmitt, 1995).

The following vignette shows how intricately problem solving is tied to emotions (*Readers Digest*, May, 1982, p. 79. Quoted from Jim Whitehead, quoted by Seymour Rosenberg in Spartanburg, SC *Herald*):

> Two hikers were walking through the woods when they suddenly confronted a giant bear. Immediately, one of the men took off his boots, pulled out a pair of track shoes and began putting them on. "What are you doing?" cried his companion. "We can't outrun that bear, even with jogging shoes." "Who cares about the bear?" the first hiker replied. "All I have to worry about is outrunning you."

Now what if the two were husband and wife, parent and child, or very close friends? The emotional investment in the relationship would change the problem to how can "we" escape or even how can "you" reach safety. A parent may suggest that the child run for it. The automatic influence of affect in defining the range of acceptable solutions to problems is evident. Westen (1993) observes that cognitive psychologists do not focus sufficiently on the motivational and emotional factors of unconscious processing of information. Affect and motivation can enhance or disrupt the degree of organization and resourcefulness in the acquisition and subsequent application of knowledge structures.

4.16.3.5 Narrative Psychology and Thematic Apperception Techniques

The analysis of narrative has been a rich source of data in studying the schema or scripts that individuals bring to current understanding of experiences. The social-cognitive approach examines the memory system by analyzing the individual's style of processing and organizing information. Memories such as self-defining experiences (Moffit & Singer, 1994) or incidents of being rejected (Baumeister, Wotman, & Stillwell, 1993) are examined to extract abstract rules and cognitive principles. A study on script formulation (Demorest & Alexander, 1992) first asked participants to generate autobiographical memories. Scripts extracted from these memories were subsequently (one month later) compared with those derived from participants' invented stories in response to affective stimuli. The scripts drawn from fictional stories were similar to those derived from autobiographical memories. These findings suggest that scripted knowledge structures are superimposed on new affective stimuli.

Schank (1990) addressed the question of how stories exchanged during the course of social conversation are accessed in memory. To understand how a person is reminded of a story to tell at the right moment in a conversation requires breaking it down into themes, plans, goals, actions, and outcomes (Schank, 1990). These elements constitute a filing system for indexing the relevant components of experiences to be stored in memory and subsequently retrieved. A higher level organization of these concepts involves a prediction, a moral, or a lesson learned. When a lesson repeatedly occurs, it becomes a type of structure that exists apart from the specific stories from which the lesson arose.

For Schank, telling a good story at the right time is a hallmark of intelligence because it represents an understanding of what is being talked about now through its connection with the lessons that the listener has stored in memory. Recalling an appropriate story to tell and understanding how it relates to what is currently being spoken about depends on how the stories are catalogued in memory. It seems logical that individuals who engage in more active and reflective classification of experience for storage into memory are also more likely to be reminded of an appropriate story at the right moment in a conversation. This framework is directly applicable to the interpretation of stories told to TAT-like stimuli. Respondents are reminded of a story by the stimulus, but the particular story details are shaped by convictions or lessons learned from experience.

One way to validate constructs is through the congruence of models drawn from various lines of research. Arnold (1962) proposed that individuals express their basic attitudes or convictions in their TAT stories, and that these convictions constitute their motivational set. Like Schank, she focused on goals, plans, actions, and outcomes as the basic ingredients of the story import or lesson learned.

The rise of narrative psychology has the potential to enrich story-telling methods of personality assessment. Analysis of narratives such as those elicited by the TAT is considered useful for clarifying the degree of complexity and organization of information processing pertaining to the synthesis of experience (Carlson & Carlson, 1984). These cognitive and perceptual processes can also be assessed through the Rorschach technique. It has been suggested that the Rorschach should combine its emphasis on perception with a focus on cognitive representation (Blatt, 1992).

4.16.4 LEVELS OF PERSONALITY ASSESSED BY PROJECTIVE AND QUESTIONNAIRE METHODS

Projective methods measure different dimensions or levels of personality than self-reports. One of the clearest ways to understand the unique contribution of projective techniques is in the context of understanding these levels. McAdams (1995) proposed three conceptually distinct levels of understanding personality: level one—traits or stylistic, habitual tendencies; level two—developmental or motivational constructs such as goals, plans, and strivings; and level three—the life narrative or the evolving story that is internalized to provide broader meaning, purpose, and cohesiveness to specific experiences. McAdams does not describe relationships among the levels, though it seems logical to assume that they are not independent since they are different perspectives (or units) on the functional whole. For Klopfer (1981), a trait itself can be understood at three levels. First, as viewed by significant others or the public image (with or without awareness); second, as viewed by the individual or conscious self-concept; and third as manifested (with or without awareness) in behaviors such as responses to projective tests. This emphasis on the source of the data is consistent with a fundamental distinction between two levels of personality, based on the perspectives of the actor and observer (Hogan, 1987). Projective tests were better able to predict peer rankings on four personality traits (publicly observed image) than the individual's self-report of those traits (McGreevy, 1962). Different techniques may be useful for elucidating different levels or dimensions of personality. However, a comprehensive understanding requires a framework for how the different levels relate to one another.

Measuring achievement motives through self-report of intentions, goals and reasons for actions (self-attributed) and inferences drawn from stories written to picture stimuli (implicit) are not equivalent (Koestner, Weinberger, & McClelland, 1991; McClelland, Koestner, & Weinberger, 1989). McClelland and colleagues propose that the general absence of significant correlations between measures of self-attributed motives derived through self-report, and implicit motives inferred from stories, cannot be ascribed to the worthlessness of projective measures nor to poorly designed questionnaires, but must be taken seriously as evidence that these are in fact different variables with the same name. Rather than seeking to remedy deficits in either type of measure, they suggest that it may be more fruitful to acknowledge them as reflecting two qualitatively different kinds of human motivation. They contend that implicit motives develop by intrinsic enjoyment generated when doing tasks or experiencing activities, and predict spontaneous goal-directed actions sustained over time even in the absence of specific social demands. However, self-attributed motives are built around explicit social incentives or demands of socializing others, and predict responses to situations structured to provide such incentives. The implicit achievement motivation does not give information about the area of life to which a person will direct efforts to succeed. Self-attributed motives, plans, and goals may express the person's conscious intentions but do not give information about the person's commitment and capacity to follow through. Under specific conditions, self-attributed motives act like implicit motives in that they energize, direct, and select behavior. However, the problem with predicting long-term behaviors from self-attributed motives is that the social incentives may not be salient enough to elicit the behavior. Implicit motives exert relatively greater influence because they drive activity that is inherently enjoyable even in the absence of specific social demands.

The acquisition of implicit and self-attributed motives such as proposed by McClelland and colleagues may relate, respectively, to the experiential and rational systems discussed earlier (Epstein, 1994). Learning in the experiential system takes place in an affective context where emotions and interests drive attention

and information processing. Additionally, the development of these two types of motives may involve different principles of learning (Meissner, 1974, 1981; Raynor & McFarlin, 1986; Sandler & Rosenblatt, 1962). One pertains to the acquisition of cognitive skills that are functional and enlarge the individual's repertoire of adaptive capacities through the teaching and reinforcement of these competencies. The other relates to the development of the inner world or the self-system that provides relative independence from external incentives. The advantage of developing this inner structure is to shift from dependence on external supports to increased reliance on self-regulatory mechanisms (Meisner, 1981). These regulatory functions increase the individual's capacity to master the environment, endure stress, and delay gratification of impulses.

The development of increasingly complex and differentiated self-regulatory structures or schema takes place at a level of organization that is different from that of acquiring behavioral patterns through reinforcement. Such development relies on the individual's integration of life experience to seek consistency within the self (Raynor & McFarlin, 1986). However, emotional, cognitive, and attentional limitations interfere with synthesis of life experiences and detract from the development of inner self-regulatory structures. For example, the domination of awareness by that which is immediately striking or concrete is not compatible with the development of long range interests, values, or aims. The primary affective impediment to the internalization of such self-regulatory values are the limitations in the capacity to sustain emotional involvement in interests that are relatively remote from personal needs (Shapiro, 1965).

Motivation can be understood in relation to two necessary ingredients: the goal or intention and the self-regulatory capacity to sustain goal directed action (Arnold, 1962; Kuhl & Beckmann, 1994). Thus, a stated goal or intention such as playing violin in a symphony orchestra may be difficult to implement and sustain, regardless of musical talent because of the rigorous independent practice required. Such sustained effort is facilitated by the enjoyment of and interest in the activity. Otherwise, careful programming of external incentives is needed.

Understanding the gaps between a person's stated goals or intentions and the individual's self-regulative resources is crucial to planning interventions. A 19-year-old college student was referred by her therapist for an evaluation because no meaningful discussions were taking place after six months of regular attendance of sessions. Sandra was in therapy at the insistence of her family because "she had no direction in life." During the initial interview with Sandra, she described her erratic schedule of eating and sleeping, her general lethargy and boredom, as well as sporadic class attendance. She reported having similar problems in high school such as difficulty getting up, missing school once a week, not studying, and cheating to get by. She stated emphatically that she was not depressed.

When asked directly about drug use, Sandra reported significant involvement starting in the fifth grade. She did not discuss this issue with her previous therapist because "no one asked me." She described the therapist as not providing enough structure during the sessions so she just talked about meaningless topics. Sandra verbalized her desire to achieve and recognized that she was not accomplishing her goals. At the same time, she admitted that she found school work aversive.

As part of the evaluation, the TAT, the Wechster Adult Intelligence Scale (WAIS-R), and Rorschach were administered among other measures. The first five TAT stories that she told are given below to illustrate the lack of connection between Sandra's goal setting and means toward their attainment. The emphasis is not on developing a comprehensive clinical picture but on presenting general conclusions about each story that are pertinent to Sandra's motivational schema:

> *Card 1.* Don't like pictures. Not all of them ... Umm ... like story—like once upon a time. This is a boy ... He's ... OK, he ... is it a violin? I don't know. Can I say two things? (Whatever you want). He's either frustrated because he can't play it or he's sad because he's not allowed to play it anymore. He wants to play it but for some reason he can't ... (stares at picture) (TO) He becomes a great (did I say that was a violin?) violin player, and he's happy. (OK, I think you have the idea.)

Sandra is immediately put off by the task and vacillates in her interpretation of the picture as she does in her own intentions. The story content is appropriate to the stimulus, but Sandra cannot determine whether the boy is experiencing internal (frustration) or external (not allowed) barriers to achievement. In any case, the boy takes no heed of these problems and becomes a great violin player, seemingly without effort. This unrealistic connection among circumstances, intentions, actions, and outcomes suggests that Sandra's schema do not include a clear grasp of the obstacles to achievement nor of the steps needed to attain goals.

> *Card 2.* God! These things are ... OK, there's ... (stares) ... This girl on her way to school, she's walking to school, and she sees this man working in the field, and that's his wife watching him. His wife's talking to him I guess, may be talking to him. He's working with his horse on the farm, and she's thinking that she doesn't want to look like that when she gets older (like the lady). She doesn't want to live on the farm. She feels sorry for that lady, for both, for the lady. So she goes to school to get an education so she's not like them.

The girl goes to school not because she really wants to learn but to avoid becoming what she fears. Sandra's schema about the importance of education does not include a positive affective quality nor the spontaneous accessibility of the effort involved. Sandra does not delve into the purposes and inner life of the characters but relies on the stimulus, on external appearances (doesn't want to look like the lady), and social convention (getting an education). Thus external structures and supports are important in guiding her reactions.

> *Card 3BM.* Don't know what this is ... umm ... guess it's a girl. A boy or girl, either sleeping, crying, or ... umm ... (frustrated) don't even know what. (repeat directions) I guess she's crying because ... OK, this doesn't make sense though. (stares as if answer is in the card) I think they're car keys. No idea. This is dumb. She's crying because ... what is that thing is ... someone stole her car. She has no way home from this place. So then she calls the police and tells them that her car is missing. Eventually they find her car, and then she's happy. Can you tell me what that is? (It can look like different things to different people.)

Sandra initially vacillates before settling on a plot. She became frustrated, stated that the task was dumb, but responded to encouragement. This tendency to blame outside factors is consistent with her reliance on the environment to regulate her behavior. The outcome of getting the car back only deals with part of the problem. Having "no way home from this place" was not addressed. Sandra's incomplete processing of information (only dealing with part of the problem, leaving out details called for in the directions) and limited resources (no internal representation of possible support; indecision and disjointed approach to narrating the story) suggest why she cannot regulate her behavior.

> *Card 4.* OK, ... this is a man and a woman who ... they either just had a fight, or he wants to go somewhere, and she doesn't want him to go. OK, she's trying to persuade him either not to go or to forgive her. And then, I guess eventually he does it, he forgives her or he doesn't—or he might go, or he'll stay. And they live happily ever after.

Again, Sandra cannot commit to one story line, and the characters' intentions or reasons for acting are not examined. As with the previous stories, she seems unable to develop her ideas clearly beyond the initial identification of the tension depicted in the card, suggesting dependence on external stimuli. Her vacillation about what's going on in the picture (noted in several picture cards) suggests that she feels uncertain about her judgment in situations that are somewhat ambiguous. As in previous stories, the ending is unrealistically positive as Sandra does not make appropriate connections among intentions, actions, and outcomes.

> *Card 5.* OK, can you make up people who aren't there? This is a lady who is coming to tell her children it's time for their nap. So she sticks her head in and says it's dinner time, and all the kids come to dinner. Everybody eats dinner, and kids go back and play. In the library, den—I think it's supposed to be the library.

Sandra's query about introducing additional characters further points to her need for external guidance. The lady intends to tell the children that it's time to take their nap but ends up calling them to dinner. This twist in the story suggests that Sandra is not monitoring story details (also seen in other stories) and that she has similar difficulty keeping track of her own intentions.

Generally, Sandra is able to recognize the conflict or tension in the pictures but has difficulty sticking to one explanation. Sometimes, she does not monitor story details sufficiently, leaving gaps or contradictory details. She does not examine feelings, thoughts, or intentions and has difficulty connecting planful, realistic actions with outcomes. She cannot develop the themes that she introduces beyond a vague and Polyanna-like story. Sandra persists in staying in school because she believes that the way to ensure her future is to get an education but is not able to get invested in the process of learning. Sandra is floundering in the unstructured campus setting because she has not developed the inner structures (implicit motives) to regulate her attention and behavior.

When she received the results of the evaluation, Sandra easily identified with the example of someone who thinks an exercise program is needed and joins an aerobics class but does not attend for one reason or another (too tired, too cold). She agreed that even the thought of doing school work is aversive, and the key was in finding ways to side-step the aversive elements, possibly through social supports. She acknowledged that by joining a sorority, she would have peers to eat with on a regular basis, and the structure would help her keep a schedule and attend class. At first, she worked on seeking the external incentives that would motivate her behavior and through these initiatives began to feel she had some choices and control over her life. Subsequently, therapy sessions focused on ways to build these structures and prompts into her daily routines to gain increasing independence.

Sandra's schema as assessed with the TAT did not include connections among goals, plans, actions, and outcomes, and these resources were not available to guide her daily behavior. The therapeutic aim was to help Sandra deal with the factors that impeded the development of these connections. She could not have verbalized these impediments in an interview or self-report measure because she was not aware of them. Likewise, the more structured Wechsler scales did not reveal the sources of her difficulty. Therefore, sole reliance on such measures would have been insufficient.

4.16.5 SELF-REPORT VS. PROJECTIVE PERSONALITY TESTING

The foregoing discussion has already demonstrated that self-reports, even if candid and accurate, may predict behavior only when social incentives are operating or when the situation is structured to prompt the response. However, it is useful to contrast questionnaire and projective devices as direct and indirect methods of personality assessment (Vane & Guarnacia, 1989). Direct methods are comprised of observations of behavior, inventories, or interviews. The indirect method utilizes relatively unstructured stimuli such as the TAT, Rorschach, or drawings. In the latter, the information sought is not readily apparent and, therefore, difficult to manipulate. Vane and Guarnacia cite the following problems with the direct method: (i) limitations of self-knowledge preclude accurate self reports; (ii) test items are open to misinterpretation; (iii) real-life situations cannot be represented by pencil and paper items; (iv) desire of the respondent to manage impressions need to be overcome by including scales to detect faking or response sets; (v) inventories yield information without clues for understanding the reason; and (vi) the relationship between traits on inventories and behaviors is complex

Multiple choice or other highly constrained response formats do not give clues about the style of organizing the response. The respondent is asked to tell versus show. Thus, no evidence exists beyond self-report to judge response parameters despite availability of norms. Finally, responses to structured questions only tap conscious phenomena. Therefore, methods such as interviews or self-report inventories provide what the individual is capable of perceiving and is willing to share. However, responses to projective testing can be evaluated by analysis of various parameters.

Both direct and indirect sources of information are useful depending on the purpose. Self-reports are susceptible to conscious or unconscious distortion and reveal only one dimension of the personality. Nevertheless, it is important to obtain information that the respondent willingly discloses. Self-reports are easily scored and amenable to traditional methods for estimating reliability and validity. Projective methods require extensive training, are time consuming, and require more cumbersome methods to establish psychometric credibility. However, their purpose is not evident to the respondent, and they provide information that is different from other sources of data. The use of both direct and indirect methods is likely to yield the most accurate picture.

4.16.6 CONTRIBUTION OF PROJECTIVE TECHNIQUES TO DIAGNOSIS AND INTERVENTION

The process of responding to the projective task can be conceptualized as an attempt to impose meaning on the stimuli presented and to comply with the instructions given. Analysis of the response reveals the individual's schema or scripts, both in terms of the structural organization of experience and the content of awareness. Appreciation that the manner in which individuals acquire, organize, and represent knowledge into cognitive schema have a fundamental influence on the way they behave, guides the use of projective techniques. Rather than looking for one-to-one correspondence between test responses and inferences, such as searching for aggressive content in projective tests, it is more fruitful to explore processes that promote aggression in particular situations. The social-information-processing model of children's social behavior (Crick & Dodge, 1994) details the social-information-processing steps that precede aggressive behavior in specific situations. According to this model, aggressive behavior can be understood in terms of how individuals encode social cues, how they interpret those cues, how they select and clarify goals, how they generate possible solutions to perceived problems or dilemmas, how they make decisions about the selection of responses, and how they execute and monitor the selected behavior. Information regarding each of these steps in social information processing can be gleaned from responses to projective techniques.

Thematic apperceptive methods can provide information about the attribution of intentions for others' reactions, the interpretation of social cues, the anticipated consequences of alternative actions, as well as about the cognitive flexibility and organization of the individual's meaning structures. Information about the execution of the behaviors is provided at two levels: how the characters follow through on intentions and resolve problems, and how the narrator develops and monitors the story. Any aggressive content that is expressed can then be better understood in light of the aforementioned processes.

Responses to the Rorschach also clarify social information processing that promotes aggression. For example, a preponderance of answers that simplifies the stimulus (e.g., high lambda) suggests that the individual may be involved in confrontational situations because of a tendency to make decisions without considering important cues (Exner, 1993). The connection between such responses to projective tasks and aggressive behavior is borne out by research suggesting that aggressive children selectively attend to and recall aggressive cues, partly because they respond prior to processing all of the available information (Dodge & Feldman, 1990).

A taxonomy of psychopathology comprised of a descriptive, atheoretical compilation of symptom categories (e.g., *Diagnostic and statistical manual of mental disorders*, 4th ed. [*DSM-IV*]) could be enhanced by a framework that relates these symptoms to basic structures of psychological organization. Within a diagnostic category such as autism, there exists a wide variability of impairment in functional capacities. Responses to projective techniques can range on a continuum from being flexibly adaptive to demonstrating various levels of impairment that can go beyond a specific diagnosis. The manner in which an individual uses the cues in the stimuli and responds in an organized way to the task demands is an index of the differentiation and organization of previously developed schema. Furthermore, responses reveal whether the person must be prompted to respond or actively restructures knowledge for adaptive use in stressful or ambiguous situations.

Horowitz (1991) envisions a future *DSM* Axis II as incorporating three interconnected components: (i) recurrent maladaptive patterns of self-regard and interpersonal behavior; (ii) person schematic characteristics explaining that behavior—this would include the developmental level of self-organization; and (iii) style of self-regulation comprised of habitual control processes leading to coping and/or defense. An example would be how a person regulates emotionality by inhibiting or facilitating the activation of schema. Horowitz's conceptualization implicitly recognizes that any Axis I diagnosis, such as attention-deficit hyperactivity disorder relates to the development of inner schema. Difficulty regulating the attentional process may interfere with the active and strategic synthesis of knowledge gained through experience. Without organized schematic structures, the individual may not be able to govern behavior according to internalized rules and standards, a prominent characteristic of individuals with attentional deficits (Barkley, 1990). Responses to open-ended projective tasks represent the convergence of all of the individual's traits on the organization of experience, including compensatory strategies that could mitigate against risk factors.

A changing emphasis within psychoanalytic theory favors the view of psychopathology as representing an impairment in the formation of psychic structures that are analogous to the schema of social and cognitive psychology. This focus has spurred diagnostic efforts to evaluate the quality of these structures (self-system, object relations). Blatt (1991) assumed that the symptomatic manifestations of various forms of psychopathology are associated with different types of impairments of cognitive-affective structures in the representational world. For example, the growing consensus about the existence of two basic subtypes of depression (empty and guilty) is based on the phenomenology of experience associated with each. One type involves exaggerated preoccupation with issues of interpersonal relatedness, feelings of depletion, dependency, helplessness, or loss. The second type occurring later developmentally is associated with issues of self-definition, autonomy, guilt, and feelings of failure. The two types of depressed individuals change in various ways in the treatment process and are differentially responsive to different forms of therapy (Blatt et al., 1988). Wilson (1988) showed how each of these forms of

depression is demonstrated differently on the Rorschach and TAT protocols. For example, stories told to TAT pictures with single characters that describe the person as empty, worthless, or helpless without introducing other characters exemplify the first type of depression. Such stories suggest that the narrator is dependent on what is immediately evident in the surroundings without being able to use inner resources to represent the support needed. The vulnerability to depression of an individual lacking such resources may be moderated by a supportive environment.

The profession is in need of a diagnostic system that identifies variables useful for the treatment process (Leve, 1995). Such a system would emphasize characteristics that interact with methods of intervention and, therefore, would serve as guidelines for their selection. The variables that Leve tentatively proposed are grouped into three categories: cognitive, environmental, and social-emotional. The cognitive and social-emotional categories include the following characteristics, among others: adequacy of causal reasoning, social-emotional reality testing, moral reasoning, ability to form and maintain close relationships, ability to identify and express emotions, as well as internal or external locus of control. Leve (1995) reasoned that cause–effect thinking is essential to the success of psychodynamic therapies since developing insight is part of the process. Cognitive treatments, however, do not require as high a level of causal reasoning because the therapist is more actively training the client either by explaining explicitly the causal connections or bypassing explanations altogether through giving behavioral instructions. Likewise, different therapies require different degrees of social intimacy for success, although all therapies have an interpersonal component. The interface of psychological variables and therapy processes need to be refined through further research. It is evident, however, that projective techniques can provide important data about these types of variables.

A fundamental consequence of schematic processing is the perpetuation of the schema in the face of conflicting evidence. Distorted schema that bias attention and information processing to conform to the original conception are related to psychopathology (Horowitz, 1991). The assessment of schema can guide psychotherapy by identifying these unconscious schematic distortions (Blatt, 1991). Bringing to awareness different schema of self (e.g., actual self, ideal self, ought self, dreaded self) and other resulted in positive therapeutic changes in both psychodynamic and cognitive therapies (Singer & Salovey, 1991).

The clinician needs a flexible repertoire of assessment techniques including projective methods because the information that each one provides is not equivalent. No method supplants or takes priority over others but contributes its part to the functional whole. The contribution of projective techniques is best addressed by looking at their place among other measures. The content and structure of responses to projective techniques can be compared with information from self and other reports. The problem-solving strategies and organization of ideas as expressed in projective responses can be compared with how these processes are expressed in more structured tasks. Each source of data is conceptualized according to its contribution to the cohesive patterns that include different levels of personality and variability in functioning under different conditions. A synthesis of these multiple layers permits more precise insight about functioning in various contexts.

Projective techniques clarify the organization and structure of the inner world. Their use is warranted whenever inner life and complex self-regulatory functions are important considerations, particularly when respondents are lacking insight or would be motivated to distort self-report. The task itself involves the utilization of previously organized knowledge structures to provide a response that is not well rehearsed. The linkages among the various elements of the response promote an understanding of the individual's organizational framework. Projective techniques add a unique dimension to the assessment by revealing the respondent's strategies for accomplishing the task and, at the same time, showing the content and organization of ideas that occupy awareness.

4.16.7 PROJECTIVE TECHNIQUES AS PERFORMANCE MEASURES OF PERSONALITY

The projective hypothesis implies that every human action and reaction bears the characteristic features of individuality (Rappaport, Gill, & Schafer, 1968). Therefore, projection is basic to every response process rather than specific to a certain set of stimuli. Essentially, any test is a performance measure designed as an analog to a life task. Projective devices are simply stimuli with known qualities that are used to elicit samples of behaviors that correlate with other behaviors.

The question has been raised about what aspects of responses to the Rorschach stimuli reflect compliance to task demands and what elements entail projection. Exner (1989) suggested that most responses are simply best-fit

answers, and that projection on the Rorschach occurs only in responses that deviate from the norm or are elaborated beyond the stimulus field. This view assumes that projection on the Rorschach is an attribute of the response and not inherent in the task. Exner explains that projection is not encouraged by the instructions which simply call for answering "what might this be" nor by the blots which are not completely ambiguous. He points out that techniques such as the TAT force projection, by asking the respondent to develop a story that goes well beyond the stimulus provided. Yet, the story-telling task also makes problem-solving demands (Holt, 1961); the narratives are expected to be organized productions rather than fantasies or random associations. Just as identification of blot contours are expected to be guided by reality constraints, so do perceptions of emotions and relationships as well as the sequence of events in stories told to TAT stimuli.

Exner claims that the operations contributing to the formulation of responses to Rorschach stimuli such as scanning, encoding, classifying, refining, evaluating, discarding, and selecting are cognitive and not projective. However, in keeping with the projective hypothesis, individuals are expected to superimpose unique styles of organization on each of these cognitive operations. The distinction between problem-solving and projection is not an either/or, but a matter of relative emphasis. All projective tasks have problem-solving elements, and all responses are influenced by individual sets or internalized schema. Those who favor the projective viewpoint insist that such inner structures are fundamental to the response process. The conceptualization of the Rorschach as a perceptual-cognitive-behavioral task has been criticized on the grounds that this approach does not give sufficient weight to the role of stimuli from the inner world in the interpretation of stimuli from the external world (Willock, 1992).

As discussed earlier, projection is an ongoing process that comes to the fore if the situation is novel. In a routine, overlearned situation, individuals will respond with familiar, scripted knowledge. However, even a highly structured situation can be misinterpreted. When ideas on the more structured cognitive tasks are associative or disorganized, these cognitive difficulties will be magnified in responding to the less structured performance tests. One youngster who demonstrated serious problems in the application of rote knowledge in a relatively structured task also showed impaired thought process on the projective tests. To the Wechsler Intelligence Scale for Children (WISC III) comprehension item, "what would you do if you saw thick smoke coming from your neighbor's home?" he responded: "Stop, drop, and roll." This response was an association cued by some of the words in the item without understanding the entire question. The fire was in the neighbor's home, yet the child produced a rote response as if he were personally experiencing the fire. He had rehearsed this scenario at school, but was unable to apply the concept even under relatively structured conditions.

The capacity to organize the inner world and deal effectively with distracting thoughts, emotions, or motives should translate into the ability to tolerate and to deal successfully with ambiguity, complexity, and apparent contradiction in a variety of situations. Therefore, the experience of ambiguity and complexity may stem not only from reality demands but also from the individual's emotions and motives (Blatt, Allison, & Feirstein, 1969). A framework is needed for conceptualizing how projective techniques such as the Rorschach and TAT shed light on cognitions in ways that differ from more structured tests. Differences and similarities in task demands of various projective techniques also need to be understood.

The response to the Rorschach task involves the matching of ambiguous stimuli with a memory trace (Exner, 1993). The goodness-of-fit or perceptual match between the blot contours and the object reported constitutes the basic element of reality testing. When telling a story about a picture portraying one or more people, what is demanded is not simply perceptual matching required by the Rorschach task (Beck, 1981) but experiential matching. The narrator searches for a relevant explanation for the picture, then marshalls possible details from memory to satisfy the instructions and meet criteria for an acceptable story. If the stimulus is highly stereotypic, then finding an exact story from memory would be sufficient to accomplish the task. However, if the narrator encounters a picture that cannot be explained by existing schema, then the individual actively constructs the response and fills in details from various schema. Schank (1990) suggests that what exists in memory is a database of partial stories or story elements rather than whole ones. When telling a story to a pictured scene, the narrator draws on this database to construct a set of events and inner processes (characters' thoughts and feelings) that captures the gist of the stimulus.

All tasks are on a continuum of how much organization is required in making meaning of the stimulus and developing the response. The more open-ended the task, the greater the demand to construct the response actively

rather than to respond in rote fashion to the stimuli. Therefore, projective techniques are particularly useful when adjustment to routine circumstances is good, but the individual is experiencing problems in less structured situations. Task by task analysis of the assessment battery permits the identification of the specific processes or competencies required for each task and the life situations to which these competencies can generalize.

A wide range of human ability lies outside the domain of standard cognitive tests. It is these competencies in the broad sense that are within the purview of projective testing. The primary advantage in regarding projective methods as performance measures of personality is that various tasks and measures can be differentiated in their demands. Generalization from one test performance to another and to real life performance then can be based on similarity of the demands. If projective techniques are treated as performance measures of personality, then it is possible to delineate task requirements, norms, and expectations about the product as seen with the Rorschach. In addition, it is possible to delineate linkages of that product with the acquisition of prior knowledge structures and with the application of these resources to similar problem-solving situations.

The term "*performance test of personality*" is proposed because it captures two key dimensions of projective techniques: (i) the problem-solving aspect of meeting specified performance expectations such as form quality on the Rorschach; and (ii) the organization of inner resources or schema brought to bear on the production. This dual approach to understanding projective techniques is shown in Table 1.

The distinction between self-report and projective measures has been described earlier. Self-reports ask the respondent to tell about the self, whereas projective measures require the performance of a task from which psychological processes are inferred. If we want to know how a person solves certain types of math problems, rather than ask the individual to report on how well he or she can solve linear algebraic equations, we would present some sample problems. Similarly, if we want to assess personality, we could obtain information through self-report by asking questions such as how flexible are you? How do you interpret cues in an unstructured situation? In contrast, we could request that the individual perform a task that demonstrates how various psychological processes are applied to meet the problem-solving demand.

Tests such as the Wechsler Scales are generally seen as measures of cognition, whereas the TAT and Rorschach techniques are tasks that are thought to reveal personality. This distinction is not entirely a function of the task but includes how the performance is evaluated. Human figure drawings, for example, have been used to estimate both cognitive level and personality functioning. The essential characteristics of tasks that measure personality, in contrast to cognition, are described below.

4.16.7.1 Total Personality

As do life situations, tasks vary in terms of the amount of structure, cues, or prompts provided for responding. As the expression goes, "It's not what you know but when you know it that matters." To be useful, knowledge structures must be spontaneously available when circumstances warrant. In the psychological test battery, personality performance tasks provide minimal guidance to meet the problem-solving demands and, thereby, permit the assessment of spontaneous strategies.

Generally, cognitive measures such as tasks on Wechsler scales do not reproduce the conditions encountered in everyday situations. Sternberg (1985) distinguished between analytic intelligence that is assessed by more structured tasks and practical intelligence. Analytic problems provide all necessary information and have a single correct solution separate from the emotional and social context. Real life problems are less clearly defined and typically call for information seeking. Note the following question on the WAIS-R: "In a movie theater, you are the first person to notice smoke and fire. What should you do?" The respondent is cued that the expectation is to do something. If this were a real situation, the person would not know that he or she was the first to see the smoke and fire. The person may not take responsibility with others around and, therefore, delay action. The individual may consider specifics of the context such as proximity to the fire, its size, or the availability of a fire extinguisher. However, the individual's emotional reaction may disrupt thinking. The actual response would involve the total personality, not merely the intellectual component.

The limitations on generalizing from responses to structured tasks to real life conditions are evident when we consider the attributes that these measures do not assess. These include being aware that a problem exists or that a task needs to be done; setting priorities and planning toward their implementation; pacing and self-monitoring; seeking or utilizing feedback; sustaining long term interest in independent activities; taking necessary risks; organizational skills; and recognizing and

Table 1 Task analysis of performance measures of personality.

Technique	*Ideal performance*	*Competencies needed*	*Inner life (how experience is organized and internally represented)*
Drawing	Symmetry, coherence, realism, match with instructions	Plan execution of drawings within allotted space and limits on drawing ability; organize details with context, handle frustration; investment in the task	Reality testing and quality of thought process; use of schema and organizational strategies; specific preoccupations or concerns
Narrative	Story matches picture and incorporates instructions, appropriate transitions and cause–effect connections; synthesis of various dimensions of experience within individual characters such as integration of inner life with external circumstances; balance among views and needs of all characters depicted; balance between excessive detail and vagueness; cohesiveness among intentions, actions, and outcomes; appropriate time perspectives	Draw on personal experience to construct a story that adequately explains the picture; modulate affects and recognize tensions depicted in the stimulus; organize, plan, and monitor details of story for cohesiveness and inner logic and for compliance with instructions; direct attention from one aspect of experience such as feelings to actions and outcomes; initiative to transcend the pictured cues to describe intentions and realistic, purposeful actions to resolve tensions; bring general abstract principles to bear on the task for optimal integration of multiple dimensions of the stimulus and coordination of inner and outer aspects of experience (thought, feeling, action, and outcome); investment in the task	Reality testing and quality of thought process; understanding social cues; strategies to organize experiences; specific preoccupations and concerns; nature of inter- and intrapersonal schema or scripts brought to interpretation of experiences and picture stimuli (experiential matching)
Inkblot	Accuracy and specificity in matching percept to form; organizing various components of the blots; balance among form and other determinants as well as between precision and vagueness in form definitions; logical and responsive communication during inquiry	Adequate investment in responding; realistic perception and organization of stimulus components; understanding of hierarchical relationships implicit in balancing form with other determinants; comfort with matching precision of percepts to relative imprecision of blots; confidence with ambiguous stimuli; systematic approach to the task	Reality testing; strategic processing of information and connecting new input with previously acquired sets or schema (perceptual matching)

Note: The task is somewhat different for each specific set of instructions and stimuli (e.g., draw-a-person in the rain vs. draw-a-person). Similarly, each Rorschach card and or TAT picture presents a unique stimulus configuration.

responding appropriately to subtle interpersonal cues. These attributes are traditionally viewed as pertaining to the personality domain and reflect the interplay of cognitive and affective processes. A response to the TAT involves the interpretation of the emotions and tensions depicted to recognize a problem, whereas a verbally communicated question or problem often contains cues to facilitate the response, and such cues may not exist in a real-life version of the scenario. Likewise, the Rorschach presents the respondent with options for interpreting the blot contours, organizing the percepts, and communicating them to the examiner.

4.16.7.2 Many Correct Solutions

Just as real-life dilemmas are amenable to diverse resolutions, personality performance measures can be approached in several ways. Therefore, many correct solutions are possible. Responses can differ according to the individual's interpretation of the stimuli and organization of the response. Variability, rather than uniformity, is the expectation with personality performance tasks. Rather than imposing specific criteria for correct responses, such tasks set a general expectation to deal with the stimulus realistically, follow instructions, and logically organize the response. In contrast, most cognitive tasks have one correct solution regardless of whether they call for rote knowledge or a more flexible application of prior knowledge. These distinctions between cognitive and personality performance tests are not absolute; various cognitive tasks share some common processes with personality performance measures. Reading comprehension, for instance, bears a resemblance to personality performance measures when previous sets influence the understanding of the text. There can be little argument about the facts presented in a passage, but individuals can justifiably differ to some extent on their inferences based on prior learning.

4.16.7.3 Generalize to Different Criteria

The distinction between typical vs. maximal performance (Cronbach, 1970) has been used to differentiate between personality and ability measures. This dichotomy contrasts responses to immediate cues and external motivating structures from performance that is regulated and maintained by the individual's long-term investment and initiative. A structured situation such as an achievement or cognitive test would generally tap maximum performance, but such performance generalizes only to similarly structured situations. Because most real life tasks provide less structure, measures of typical performance are also needed in a comprehensive assessment. Correlations are low between measures tapping maximal and typical performance (Sackett, Zedeck, & Fogli, 1988). Therefore, each provides unique information.

The following conditions were suggested by Sackett and colleagues as generally yielding estimates of maximum performance: (i) there is a heightened level of effort and attention because the task is seen as important; (ii) expectations and performance standards are clear; and (iii) the observation takes place over a relatively short time where the individual can exhibit an uncharacteristic spurt of effort that could not be sustained over the long haul. In contrast, the characteristics of measures of typical performance are as follows: (i) individuals are unaware that they are being observed or evaluated, so they are not trying deliberately to perform to the best of their ability; (ii) performance is monitored over a long period of time; (iii) the performance tasks require skills that have to be learned through continuous past efforts—if the task is highly complex, the individual has to bring a great deal of past learning (typical performance) to the current effort; and (iv) performance guidelines are not clear and, therefore, individuals impose their characteristic way of organizing and dealing with the situation.

Performance measures of personality meet several criteria for assessing typical performance. Prior knowledge (schema) is superimposed on current task demands; individuals are unaware of what aspects of their performance are evaluated; and in the absence of structure, they are required to organize and plan their response according to their typical mode of functioning. Understanding different requirements of different tasks helps explain variations in the assessment of similar constructs with different methods. Johnston and Holzman (1979) designated indices of thought disorder in the WAIS and the Rorschach which they tested with a sample of schizophrenics, their parents, and controls. The IQ score and the Thought Disorder Index (TDI) derived from the WAIS were negatively correlated such that the higher the IQ score, the lower the TDI. In contrast, IQ scores and the TDI derived from the Rorschach were uncorrelated. The authors conclude that the tasks make different demands. The WAIS calls for "habitual reactions and the social frame of reference is clear" (p. 61). The social expectations on the Rorschach are less obvious. Therefore, the task sets different requirements and has different implications for the assessment of impaired thought process. In their sample of schizophrenics, those who were more intelligent were able to limit expression of disordered thinking on the WAIS more easily than on the Rorschach, in part, because WAIS items can be answered with overlearned responses. The feeling of ambiguity versus security determines efficiency of functioning on problem-solving tasks and in real life. Individuals who exhibit thought problems only on unstructured tasks may be more capable of functioning with the supports of structures and clear social cues.

A criticism of laboratory research has been that results do not generalize to nonlaboratory situations (Fromkin & Streufert, 1976; Snow, 1974). This is the case because precise linkages between processes engaged by laboratory activities and other life domains are not specified. Likewise, predictions from test performance to real life adjustment can be accurate only if they are functionally similar. By understanding patterns of strength and weakness across various tasks, the professional can point to areas where the individual can and cannot function adaptively.

4.16.7.4 Differences in Conditions of Learning and Performance

Responses to performance measures of personality, as most tasks, require previously organized knowledge and strategies. However, we distinguish between learning that is promoted by direct teaching (e.g., lecture, textbook) from learning mediated by the individual's synthesis of experience (Epstein, 1994). Personality performance tasks are guided more by self-regulated learning than by formal education. Again, this distinction in the conditions of learning does not apply in a dichotomous fashion to personality versus cognitive measures. For example, an individual's general fund of information is a joint function of direct teaching and the individual's interest and active, effortful processing of information. Depending on the individual, knowledge is acquired through some combination of direct teaching and self-regulated synthesis of experience.

Task requirements also differ on the basis of the conditions of performance. These conditions pertain to the spontaneous versus cued accessibility of prior knowledge and degree of organization required to produce the response. Cognitive measures that assess general fund of information often elicit previously acquired knowledge in piecemeal fashion (highly structured). Personality performance measures such as the TAT and Rorschach set conditions of performance that demand spontaneous accessibility to prior knowledge to interpret the stimulus and formulate the response according to the directions. The more open-ended the response, the greater the need for self-regulated strategies to organize the product.

Performance measures of personality maximize the imprint of organization so that the principles by which experiences are structured and the inner organization of the personality are revealed. A comprehensive assessment battery provides tasks that vary on the continuum of structure provided. Given such variation, it is possible to relate competencies required in each task to performance in life situations requiring similar competencies. This linkage is accomplished by understanding how processes exhibited during test performance are carried over to everyday functioning in various situations and incentive conditions.

4.16.8 SPECIFIC PROJECTIVE TECHNIQUES

Three basic aspects of projective techniques have been recognized (Rabin, 1981). First, the task entails the presentation of an ambiguous set of stimuli and a request to give an open-ended response. The Rorschach and TAT techniques include both of these stimulus and response attributes. Therefore, the focus of interpretation is on the perception of the stimuli presented and the organization of the response. Drawing techniques also demand open-ended responses but usually do not provide a stimulus. Second, the response is shaped by processes that are outside of conscious awareness. An important factor making the response less amenable to conscious manipulation (faking) is that the respondent does not comprehend the meaning of the answers given. Third is the complexity of the interpretive process. Each of these components is briefly addressed below.

4.16.8.1 Stimulus

In general, projective techniques present stimuli that are amenable to various interpretations, and instructions that can be addressed in a variety of ways. The degree of ambiguity has been a prime consideration, although other features of stimuli are also important determinants of the response. A systematic accounting of how the respondent uses stimuli is the essence of the Rorschach technique and must be given greater weight in thematic apperceptive methods (Henry, 1956; Teglasi, 1993).

4.16.8.2 Response

All projective tasks require the individual to draw on internal images, ideas, and relationships to create a response. The respondent must

dredge forth past experiences, direct or vicarious, and organize them to meet the task demands. The greater the stimulus ambiguity and the more open-ended the response, the greater the reliance on the organizational structures of the personality rather than on rote knowledge. Yet, the stimulus must have sufficient structure to permit evaluation of the plausibility of the respondent's interpretation. The projective task demand is an analogue of other unstructured tasks and situations where available cues are subject to interpretation. The manner in which the individual interprets the stimuli and organizes the response shows how they will respond under similar conditions (Bellak, 1975, 1993).

Responses to projective tasks involve complex, interrelated processes that have conscious and unconscious components. These include the interplay of cognition-emotion-action tendencies that coordinate perceptions of the outward world with experience of the inner. Projective techniques may reveal aspects of emotion, motivation, and cognition that a person may not wish to expose. These unconscious aspects of responses may relate to issues of invasion of privacy. Faking, malingering, or defensiveness are problematic for any form of assessment but are assumed to be less so with projective techniques. However, the issue of fakability of projective tests has been inadequately investigated (Rogers, 1997).

4.16.8.3 Task Demand

The stimuli presented together with the instructions set the task demands (Teglasi, 1993). Projective techniques impose task demands that cannot be met with a simple response such as a request for specific information. They require respondents to apply what they know to produce a story or a drawing or to identify an object that may fit the ambiguous contours of an inkblot. Various projective tasks have features in common yet differ in important ways. Projective methods have been understood as problem-solving tasks with designated performance expectations as well as measures of personality with emphasis on individual variation.

4.16.8.4 Interpretation

The interpretive task is complex even when the scoring categories and interpretive guidelines are straightforward as in the Comprehensive System for the Rorschach. Despite the relatively clear guidelines and availability of norms (not to mention computerized reports), the Rorschach requires the examiner's trained inference along with the more objective coding. Interpretation is in keeping with the clinician's theoretical framework and understanding of the task demand. Therefore, the interpretation of responses to a projective test cannot be more satisfactory than the adequacy of the theory informing the interpretation of the evidence and the examiner's skill in evaluating that evidence.

The complexity of the interpreter's job is maintained because conclusions rest on the understanding of meaningful patterns rather than isolated response elements. Personality performance tests yield products where the whole is more than the sum of the parts. The evaluation of the response must account for the organization and cohesiveness of the different components. Therefore, various units abstracted from the whole cannot be treated in a piecemeal fashion. The more open-ended the response, the more amenable to separate analysis of structure, form, and style. Although these aspects of the response can be conceptually separated, their interpretive value lies in their relationship to the task demand and to each other. Content, for example, cannot be properly understood apart from the manner in which it is organized. Finally, the examiner must be aware of connections between empirical findings and theory to avoid speculation. Yet, empirical support of theory is not always conceptualized in ways that are useful in making decisions about one individual.

The sign approach attempts to provide empirical evidence for interpretations through the identification of features that occur most frequently in specified clinical populations. However, using a list of signs in an atheoretical cook book fashion is inadequate because a particular sign derives its meaning from the context of other responses. For example, in a TAT protocol, a stereotyped approach or meticulous listing of details in story telling may be viewed as resistance or as representing the respondent's best efforts. Likewise, concern with minutiae that most respondents disregard may be viewed as an index of hypervigilance or of concrete functioning. The appropriate interpretation depends on the pattern of responses within the story telling task and across tests in the battery.

Inferences are drawn from the clients' behavior during the evaluation as well as from the product. Any changes in demeanor or emotional reactions in response to the various tasks are noted as are spontaneous comments, time elapsed, expressions of uncertainty about performance, or attempts to seek structure. Aspects of the response process are also considered. These refer to the manner of working, compliance with instructions, sequence of ideas (e.g., planful, trial and error, organized, haphazard), as well as to dysfluencies, pauses, or hesitations. The product itself is amenable to analysis in relation to the process of the response, content, and structure. The formal or structural aspects of the product, of course, reflect the response process.

Next is a brief review of three major projective techniques in terms of the following elements: (i) task demands that include the stimuli and instructions; (ii) response process; and (iii) general interpretive approaches and issues.

4.16.9 THEMATIC APPERCEPTIVE TECHNIQUES

Apperception tests generally use pictures and standard instructions to elicit stories. The TAT (Morgan & Murray, 1935) is the most popular of these. Although introduced as a method to assess a particular theory of personality (Murray, 1938), this technique did not remain wedded to the theory of personality from which it sprang. The thematic approach described in the standard manual (Murray, 1943) was amenable to use with a wide range of scoring methods and has been adapted for clinical and research purposes. Spin-offs of thematic apperceptive approaches have used different picture stimuli and diverse scoring approaches. A variety of nonclinical coding systems for content analysis of the TAT appears in an edited volume by Smith (1992a). More clinically oriented systems also abound (e.g., Bellak, 1975; Cramer, 1996; Henry, 1956; Karon, 1981; Rappaport, Gill, & Schafer, 1968; Teglasi, 1993; Tomkins, 1947; Westen, 1991; Wyatt, 1947). Typically, clinicians have used the TAT by applying broad units of inference in an idiographic, qualitative manner. In contrast, researchers have scored TAT responses for various personality characteristics using specific and narrow scoring criteria. As of this date, there is no widely agreed upon scoring system for the clinical use of thematic methods. However, it is generally acknowledged that classification into diagnostic categories is not a chief purpose of the TAT.

4.16.9.1 Response

When given TAT-like pictures, the scene sets the topic. The story telling directions call for a description of what is happening in the picture, what happened before, what people are thinking, how they're feeling, and how everything turns out at the end. These instructions require the narrator to attribute thoughts, feelings, and motives to the characters depicted and embed these inner aspects of experience into an appropriate time frame and sequence of events that coordinate inner life with appropriate actions and outcomes.

4.16.9.2 Stimuli

Picture stimuli play a major role in determining the story content (Kenny, 1964; Murstein, 1965), and reviews of frequency of themes specific to TAT stimuli are available (Bellak, 1975; Henry, 1956; Holt, 1978; Murstein, 1968; Stein, 1955). Ambiguity was a primary consideration in designing the TAT stimuli (Murray, 1938). However, the degree of ambiguity varies within the set of TAT cards. The manner of defining ambiguity has ranged from judges' estimates of the number of interpretations that can apply to each card (Kenny & Bijou, 1953) to actual degree of variability in responses (Campus, 1976). Many of the TAT pictures clearly show who the characters are, what they are doing, and the emotions they are experiencing (Murstein, 1965). Yet, they permit a great deal of variation in the style of expressing similar themes. While stimulus ambiguity is essential to the projective hypothesis, other issues pertinent to stimuli such as similarity of main character(s) to the narrator, emotional tone, complexity, and latent meaning have been researched (see review by Teglasi, 1993).

Broad conclusions about the advantages of various degrees of stimulus ambiguity cannot be drawn without specifying the nature of the scoring system, population, and purpose. Low ambiguity has been favored in the assessment of specific motives (Singer, 1981). However, cards with high ambiguity have been preferred to measure the relative strength of two motives (Atkinson, 1992). When studying hostility, it was found that aggressive intent was most effectively measured by a picture with low relevance for hostility, whereas guilt over hostility was best measured with a picture of high relevance for hostility (Saltz & Epstein, 1963). The authors concluded that pictures with low relevance for unacceptable behavior such as hostility measure drive toward its expression, and those with high relevance measure inhibition or guilt about its expression.

Three degrees of ambiguity within the TAT set were estimated by a group of judges—high, medium, and low (Kenny & Bijou, 1953). The richness of personality content obtained varied as a function of card ambiguity. Cards in the medium ambiguous set yielded stories with the most personality information. The authors discussed two dimensions of ambiguity in card stimuli: (i) the number of cue constellations

available to guide the response; and (ii) the definitiveness of the cues available. Sets of stimuli with graded levels of ambiguity permit the evaluation of responses to situations along a continuum of structure.

Studies of stimulus ambiguity are inconclusive because investigators compare various degrees of ambiguity that are not on the same points on the ambiguity dimension. Furthermore, conclusions about stimulus variables have generally been drawn regarding their adequacy to elicit content germane to the assessment of specific motives or relatively narrow psychological processes rather than their effectiveness to assess formal qualities of organization that reflect the respondent's capacity to deal with the task demands.

For clinical purposes, stimuli must permit the simultaneous evaluation of multiple psychological processes and analysis of both structural and content properties. Some clinicians argue that stimuli are relatively unimportant as long as they are not highly structured (Arnold, 1962; Karon, 1981). Criteria for picture selection within the TAT set are available (Bellak, 1975; Birney, 1958; Henry, 1956; Teglasi, 1993). Haynes and Peltier's (1985) survey of clinicians in juvenile forensic settings indicated that the mean number of cards used was 10.25. The most frequently administered TAT cards were similar to those deemed as most productive in previous studies (Cooper, 1981; Hartman, 1970). The TAT pictures have been criticized for being out of date in relation to clothes and hairstyle (Henry, 1956; Murstein, 1968) and for their predominantly negative tone (Ritzler, Sharkey, & Chudy, 1980). These concerns are mitigated by findings that the stimuli effectively permit expression of convictions and conceptualizations of affect. The negative tone presents unfinished business or a dilemma to be resolved and provides the opportunity to observe how the respondent interprets the tensions depicted in the picture. Furthermore, the negative scenes make it possible to observe how the narrator moves beyond the sadness or conflict to a more positive resolution by noting appropriateness of transitions between the negative state of affairs and positive outcomes. It has been suggested that pictures representing relatively universal social situations that most people encounter in their lives are suitable across various age and subcultural groups (Veroff, 1992).

4.16.9.3 Interpretation

Karon (1981) suggests that only a complex scoring system can be clinically useful. At the same time, "an adequate scoring system is so complex that no one has time to score the protocol" (pp. 94–95). Simple scoring systems are inadequate because they "throw away most of the information in the process of scoring, and hence turn out not to be clinically useful" (p. 95). This point is well taken. A multiplicity of thematic methods are available that provide carefully described scoring systems to assess relatively narrow aspects of personality. Although these systems have demonstrated a high degree of accuracy and reliability, their narrowness limits their clinical utility. Karon's remedy is to rely on clinical judgment in a sentence by sentence interpretation of the protocol, much like a clinical interpretation of an open-ended interview. This approach is compatible with the view that projective techniques are not tests, but are clinical tools that rely on the skill of the practitioner (Anastasi, 1976). An alternative is to agree on key psychological processes and guidelines for their measurement that may be time consuming to learn but less cumbersome to use once mastered.

Historically, there have been two ways to designate units of analysis for TAT stories: product centered versus narrator (person) centered. The former method focuses on qualities of stories that differentiate between groups; the latter approach emphasizes the psychological constructs to which attributes of stories relate. The narrator-centered approach permits the hierarchical organization of elements identified in stories in reference to broader features of the personality. For example, Rappaport, Gill, and Schafer (1968) organized story qualities according to affective lability and looked for content shaped primarily by affective responses to the picture. Holt (1958) looked at clarity of thought (vagueness, overgeneralization, disjointedness of organization) and emotional inappropriateness (arbitrary turn of events, forced endings). With such a focus on psychological processes, patterns of story elements can be organized in relation to relevant constructs.

Holt and Luborsky also emphasized adequacy of hero. This variable, however, is not a psychological process but a quality of the story considered important because it correlated with supervisor's ratings of overall competence during psychiatric residency training. Adequacy of the hero was subsequently viewed by Bellak (1975) as an index of the narrator's competency. However, this interpretation may not hold if other characters depicted in the stimulus or introduced in the story are helpless. This conclusion is also problematic if the narrator has not accurately incorporated the cues provided in the stimulus. A narrator-centered approach to this variable would focus on various psychological processes indicative of the story teller's competency. Three points are relevant here. First, no part of the story can be interpreted separately from the others. Second, the emphasis must always be on the narrator rather than the narrative. Third, the demands set by stimulus properties must be considered.

Emphasis on qualities of the narrator rather than of the product permits consideration that these qualities can be expressed in different ways. Therefore, the interpreter does not seek one-to-one correspondences but focuses on the fit between the pattern of responses and the psychological processes to which they pertain. For example, the assessment of a construct such as cognitive integration with thematic apperceptive methods may be accomplished by exploring the manner of interpreting stimuli or the manner of coordinating different dimensions of experience such as thoughts, feelings, and actions. More complex units, such as self-regulation, subsume cognitive integration as well as affective-motivational processes (Teglasi, 1993). Such a focus on the psychological processes of the narrator avoids the sign approach and promotes an understanding of the constructs being measured. This occurs by organizing story qualities in terms of a clear conceptualization of their relationship to relevant psychological variables. Conceptualizations of the psychological variables can undergo continuous refinement in keeping with research across various subdisciplines.

Form and content in the TAT have been distinguished (Henry, 1956; Holt, 1958; Teglasi, 1993). Formal features focus on generalized properties of content such as the organization and coherence of the response. These formal units of analysis refer to how the details of the content relate to each other, to the evoking stimulus, and to the directions given. For example, inferences can be made from TAT stories about striving for long range goals (independently of the specific goals or concerns) based on connections among characters' purposes, actions, and anticipated outcomes. Internal attribution of feelings is a formal quality because it need not refer to any specific feeling. These formal qualities of the story provide information about the cohesiveness and reality base of the narrator's schema rather than about specific concerns. They are akin to procedural knowledge within schema theory and should be evident across stimulus cards within a task and even across various projective tasks. Formal features of stories lead to inferences about the structure of personality such as cognitive integration, affect maturity, or self-regulation. Given that picture stimuli evoke specific content, it is these generalized or formal properties of content that can reliably reveal clinically important psychological processes.

A focus on the formal elements involves the following general principles of TAT interpretation.

(i) An organized narrative, not a patchwork of associations, is expected. The story is, therefore, evaluated in terms of how the narrator meets problem-solving task demands. The individual's problem-solving approach can be examined from two vantage points. The first focuses on how the story is told in relation to the stimulus and instructions. Accordingly, the narrative is evaluated in terms of accuracy in capturing the tensions depicted in the stimuli, compliance with instructions, and the logical and realistic unfolding of events. The second emphasizes how the characters are described in reference to the stimulus cues, the cohesiveness of inner states, actions and outcomes, as well as the manner in which characters define and resolve the dilemmas set before them in the stimulus.

(ii) Units of inference based on interconnections of elements such as the links between causes and effects or actions and outcomes pertain to formal characteristics of content. The content is not taken at face value. Instead, units of inference are based on understanding of psychological processes and task demands. The instructions call for inclusion of various levels or dimensions of experience such as thoughts, feelings, and actions. These human tendencies to think, feel, perceive, and act are interrelated, and interpretive meaning is derived from an understanding of their patterns. However, the narrator's understanding and the clinician's framework for conceptualizing such relationships must be clearly distinct (Cramer, 1996). The professional determines the implications of the narrator's understanding of the linkages among sequences of events, thoughts, feelings, behaviors, and outcomes.

(iii) The interpretation seeks patterns that elucidate the schematic structure connecting the various story elements (e.g., the manner in which self and other are differentiated and expectations about causal sequences). These schema are products of the synthesis of past experience and provide the templates for organizing current experiences.

4.16.10 RORSCHACH TECHNIQUE

The Rorschach has achieved wide acceptance due in large part to the development of the comprehensive system (Exner, 1993), an integration of previously established methods. The

availability of normative data for discrete coding variables establishes psychometric credibility. In addition, the comprehensive system provides interpretive strategies for synthesizing the complex data and encourages the interpreter to move back and forth between the formal data of the structural summary and the content and language of the response.

The Rorschach administration is conducted in two phases. First is the association phase during which the respondent is shown each of 10 inkblots and asked "What might this be?" The inquiry phase begins after the respondent reports what is seen on each of the cards. The examiner guides the respondent to clarify the determinants of each response to permit accurate coding. Although the Rorschach is described as a cognitive-perceptual task (Beck, 1981), what is interpreted is what the respondent verbalized. Hence, the examiner must be well trained in coding and administrative procedures to avoid the many potential pitfalls of conducting an inquiry and to maintain the delicate balance between seeking sufficient clarification and promoting response sets by pressing too far.

4.16.10.1 Stimulus Features and Response Parameters

The structural approach to interpreting responses to Rorschach cards creates a close link between the stimulus qualities and parameters of the responses that are coded. Stimuli vary in the degree to which they are solid or broken, colorful or achromatic. They also display variation in shading, empty (white) spaces, and contours that are sufficiently familiar to evoke popular responses. Reported perceptions are coded according to patterns of attending to stimulus qualities including choices of blot areas and use of various features of the stimuli such as form, color, shading, hue, white space, card symmetry, or any combination. Also important are the accuracy of the match between the reported percept and selected blot contours and the relationships among the percepts (organization). Each code assigned to a response represents a particular psychological process. However, specific response elements such as choice of blot areas are interpreted in relation to other variables such as degree of organization in relating various blot areas. A response involving the synthesis of parts into larger units reflects more complex thinking than a vague, undifferentiated holistic perception of the card.

Perception of movement is not attributable to a quality of the stimulus which is static, but an imposition of the perceiver. It has been argued that phenomena such as movement are not perceptions (because blots are static, not moving) but mental constructions based on perceptual experiences (Blatt, 1992). In general, movement responses attempt to put greater specificity on form qualities. When such responses involve human activity, they often represent the individual's use of inner resources to modify perceptions of the external world. However, this interpretation changes if these efforts are combined with inaccurate form identification. If the forms identified differ greatly from those reported by others in a culture or society, then the individual is likely to make other people uncomfortable (without necessarily knowing why) by engaging in behaviors that depart from their expectations. Use of white space as figure-ground reversal is an analog to other figure-ground reversals such as an oppositional stance toward the world. Conventionality is interpreted according to the extent to which thought process is like others or different. The shading is a relatively subtle aspect of the Rorschach stimulus so that not all respondents report them as determinants (although they may perceive them). Such responses require a certain level of sensitivity or perceptiveness to verbalize.

The style of communication and nuances of language are important considerations (Smith, 1994). Therefore, characteristics of verbalizations such as redundancies or logical inconsistencies are noted. Even commonly reported responses can be expressed in unique ways. Percepts on the Rorschach are at times given without commitment or reflection, and the examinee might in such instances be reluctant to engage in the inquiry just as he or she was removed from the response process. These stylistic qualities of the response process provide a context for the interpretation of the variables in the structural summary.

Content given is also considered. For example, those who give a broad range of content have a greater variety of interests, and those who provide many human associations are exhibiting a broader interest in people. However, what is emphasized in coding responses is the manner in which the respondent uses the stimulus.

The Holtzman Inkblot Technique (HIT) (Holtzman, Thorpe, Swartz, & Heron, 1961) was designed to overcome what had been perceived as psychometric limitations of the Rorschach. Two parallel sets of inkblots were constructed, each set containing 45 inkblots plus two practice blots that are identical for both sets. The HIT was designed to differ from the Rorschach in several ways besides the number of inkblots (Holtzman, 1981). Two important differences are that the respondent is instructed to give only one response per card rather than leaving it open and that a short and restricted inquiry is given after each response. This overcomes criticisms of the Rorschach arising from the variations in the style of inquiry and from the widely varying number of responses obtained which make it difficult to use norms. However, abandoning the principles basic to the Rorschach also has disadvantages. One cannot observe whether the respondent can see something different in an inkblot once it has been identified in a particular way (flexibility), and the number of responses *per se* is an important consideration. Sequential analysis of multiple responses to the same stimulus is not possible in the Holtzman format.

Clearly, in devising variations on techniques, there are trade-offs that need systematic attention. Given such acknowledgment, clinicians can make appropriate choices.

4.16.10.2 Interpretation

The process of interpreting the Rorschach has distinct phases. First, the examiner codes the responses and organizes them into patterns by calculating the variables in the structural summary. These variables are then compared with norms to designate deviations from expected patterns. A strategy for interpreting the protocol is selected based on constellations of responses that depart from the norms. The final step involves the synthesis of the norm referenced response patterns according to the psychological process associated with them. This step also involves analysis of the details of content and of the quality of the verbalized responses. The comprehensive system provides clear guidelines for each step of the interpretation process (Exner, 1993), but the examiner integrates information from various sources to refine conclusions and formulate recommendations.

The comprehensive system began with an almost exclusive emphasis on formal, structural characteristics emphasizing the perceptual-cognitive nature of the task to match blot areas with objects (Weiner, 1994). However, language and other associations are considered important (Rappaport, Gill, & Schafer, 1968) and have become increasingly incorporated into the interpretive procedure (Exner, 1993). Smith (1994) cautions against acceptance of the Rorschach as a test with one correct method of interpretation. An exclusive focus on a single approach fails to acknowledge the potential contribution of those with a different perspective. Content interpretation including recent attempts at codings of object relations (Lerner, 1992; Stricker & Healy, 1990) is difficult to reconcile with the emphasis on structural variables. Historically, these two types of data have been viewed through different theoretical perspectives. An alternative suggested by Weiner (1994) is to focus on distinctions within personality rather than on classes of data such as form versus content.

4.16.11 DRAWING TECHNIQUES

Drawing tasks do not provide a stimulus but do exhibit the dual nature of projective devices as problem-solving tasks that reflect internal representations. Drawing a person or a family requires the translation of a three-dimensional memory image into a two-dimensional graphic representation within the constraints of the respondent's artistic ability. This conversion of three dimensions into two and compliance with instructions are the chief problem-solving task demands. Performance of the drawing is affected by the conceptualization of the object drawn, motor execution, attention to detail and spatial relationships, as well as planning and organizing the production. Like other projective tasks, the drawing can be evaluated according to its structure (proportion, elaboration, consistency of detail), content (what is drawn), and style (line pressure, size, placement). Machover (1949) suggested that structural or stylistic elements of size, placement, quality of line, positioning, symmetry, elaboration, or shading are more reliable than contents such as body parts and clothing. Performance standards for drawings can be set such as respect for inside and outside boundaries of the persons drawn and coherence and balance among details. A stylistic quality such as perfectionism and frequent erasures can be compared to the eventual quality of the product. The sheet of paper sets the boundaries or limits for the drawing. Very small or very large drawings suggest that the individual has difficulty setting boundaries in relation to the environment. However, other influences such as impaired psychomotor functioning (e.g., arthritis) in older clients (Kahana, 1978) must be acknowledged. Furthermore, processes such as lack of planning contribute to the use of available space. The individual may draw a head that is too large (lack of planning) so that the drawing either is out of proportion or cannot be completed on the page.

Drawings have been used to estimate intellectual functioning and neurological status as well as to assess personality. The draw-a-person Test (DAP) widely used as a projective

technique to assess personality (Machover, 1949) is also used to measure mental development (Goodenough, 1926). Its usefulness as an indicator of concept formation (intelligence) in the developing child was based on how closely the drawing approximated life-like proportions and details. Basically, the degree of realism was the criterion reflecting the child's conceptualization of the outward world through correct depiction of space and perspective. The Goodenough (1926) Draw-A-Man Test and the Harris (1963) revision of some criteria included items that appeared more frequently with increasing age. However, even as emphasizing age related trends, Goodenough (1926) noted the following individual variations in drawing that were not linked to development: (i) detail dominated but with few ideas; (ii) unique depictions that seem comprehensible only to the drawer; (iii) suggestive of a flight of ideas; and (iv) contradictory combination of primitive and mature characteristics.

Koppitz (1968) also differentiated qualities of children's drawings that were not a function of age. These aspects of drawings were stylistic: (i) qualitative aspects of integration, symmetry, and shading; (ii) presence of unexpected characteristics; and (iii) absence of expected characteristics beyond various ages. These qualities were viewed as manifestations of concept formation that were impacted by motivational and emotional aspects of the personality rather than chronological development. Drawing tasks are frequently given to adults to assess neurological status based on the observation that brain damage interferes with the integration of spatial, perceptual, and motor responses needed to execute drawings (Swindell, Holland, Fromm, & Greenhouse, 1988; Mendez, Ala, & Underwood, 1992).

Currently, the most frequently used projective drawing technique is the Draw-A-Person (DAP) as developed by Machover (1949) and expanded by others (Hammer, 1958; Handler, 1985; Koppitz, 1968, 1984; Urban, 1963). Also popular are the kinetic family drawing (KFD) technique (Burns, 1987; Burns & Kaufman, 1970, 1972); and House–Tree–Person (H–T–P) drawing task (Buck, 1948, 1987).

4.16.11.1 Interpretation

The dichotomy between person centered and product centered interpretation mentioned earlier is relevant here. Person centered approaches organize aspects of the production according to relevant psychological processes of the client. The interpreter looks for coherence in form, content, and style of the production with the hypothesized psychological variables. The professional also notes consistency with drawing instructions, nature of verbalizations about the figures drawn, and the amount of effort expended. Global interpretations of interrelated parts are more valid than isolated interpretation of specific details. The sign approach to validating drawings has been criticized (Kahill, 1984; Roback, 1968; Swensen, 1968) because there are multiple possible interpretations for any quality, depending on the overall pattern. For example, size and detail of drawings can signal preoccupation, differential valuing, conflict (Fisher, 1986) and/or quality of planning and organization. Similarly, qualities such as shading or erasures are said to signal anxiety. However, the nature of the shading is important since this feature may enhance artistic quality and not reflect anxiety. It should also be noted that people differ widely in how they cope with anxiety. Some anxious individuals will race through the task without shading or erasing. If the focus is on the psychological processes of the individual, then the configuration of response elements or signs would be expected to cohere around that process.

It has been suggested that the clinician ask what a particular sign could mean rather than what it does mean (Handler, 1985). A critical review of the literature on the KFD technique (Handler & Habenicht, 1994) emphasized the need to study more holistic, integrative approaches to the KFD rather than the interpretation of a series of single signs. The authors argue for the importance of focusing research on the interpretive approach of the clinician using the KFD rather than on the technique itself. The use of drawings for multiple purposes (e.g., concept formation, neurological status, or personality) demonstrates the multidimensional aspects of projective tasks which permit simultaneous evaluation of responses from multiple perspectives to reveal various interrelated facets of functioning. Broad performance expectations pertaining to structural features of the product can be delineated. As with other performance measures of personality, individuals are free to vary their approach to meeting these general problem-solving expectations. One caveat is that artistic quality of the drawings appears to influence clinician's interpretation (Feher, VandeCreek, & Teglasi, 1983).

4.16.12 CASE ILLUSTRATION

On the surface, the three projective techniques reviewed seem very different. Yet, they reveal similar information when compared in terms of structural qualities. The overlap between the TAT and Rorschach is illustrated by comparing formal aspects of Carl's TAT stories with conclusions drawn from the Rorschach. Carl, a 19-year-old man, had stopped attending school on a regular basis after the seventh grade and left school altogether during the 10th grade. At the time of this evaluation, Carl was incarcerated and attempting to continue his education in prison. He was doing poorly in his classes and was referred to determine if he would qualify for special education services. His WAIS-R IQ score was low average. The intent here is not to develop a comprehensive clinical picture but to point out the overlaps between the two projective techniques. Table 2 shows the consistency of the TAT variables across cards and also displays the relevant Rorschach data.

Card 1. Little boy thinking, he's thinking how. I guess he's thinking a way how to work the violin. (Before?) He don't know how to work the violin. (TO?) Turns out he's still sitting there thinking.

Table 2 Corresponding aspects of Carl's TAT stories and Rorschach variables.

	Card									
TAT variables	1	2	3	4	5	6BM	8BM	12M	13MF	13B
Imprecise accounting of the stimulus		X	X	X		X	X		X	
Vague, concrete, or stereotypic story	X	X	X	X	X	X	X	X	X	X
Requires more than one query	X	X	X	X	X		X	X	X	X
Story ending is concretely tied to the stimulus	X	X	X	X	X	X	X	X	X	X
Relationship among characters are unclear, stereotypic, or unstated	NA	X	NA	X	NA	X	X	X	X	NA
Sense of helplessness (e.g., inaction, when it is warranted; lack of initiative or inertia)	X		X	X	X	X		X	X	X
Inner life such as intentions for actions or feelings is not sufficiently elaborated	X	X	X	X	X	X	X	X	X	X
Implausible sequence of events (e.g., cause–effect, timing, coherence of action with purposes)	X	X	X	X	X	X	X	X	X	X
Focus is dominated by the immediate circumstance or consideration	X	X	X	X	X	X	X	X	X	X
Insufficient integration of detail to comply with instructions	X	X	X	X	X	X	X	X	X	X

Rorschach variables
Resources and control
EB = 0:0; EA = 0; eb = 7:0; es = 7; *D = –2; *Adj. D = –2
Affect
FC:CF + C = 0:0; Pure C = 0; *Afr = .2143; S = 0; *Blends: R = 0:17; CP = 0
Interpersonal
* COP = 0; AG = 0; Food = 0; Isolate/R = .1764;* H:(H) + (Hd) + Hd = 0:0 *H + A:Hd + Ad = 13:0
Ideation
a:p = 4:3; Sum 6 = 0; M^a:M^p = 0:0; Lvl 2 = 0; 2 Ab + (Art + Ay) = 0; W Sum 6 = 0; M– = 0; M none = 0
Mediation
*P = 4 *X + % = .29; *F + % = .30; *X–% = .4117; S–% = 0; *Xu% = .29
Processing
Zf = 12; *Zd = –.8.5; *W:D:Dd = 13:1:3; *W:M = 13:0 *DQ + = 3; *DQv = 4

* Deviates from normative expectation; R = 17; Lambda = 1.428; PSV = 3; Positive indices: SCZI; CDI.

The boy in the story can only keep sitting there thinking how to work the violin. He does not have sufficient resources to deal with the task or to find alternatives. Likewise, the narrator cannot go beyond the picture cues provided (concrete) to produce an ending or to describe purposes or deliberate actions. When prompted to tell what happened before, the response (didn't know how to work the violin) stayed within the moment without considering larger purposes. The fact that the narrator does not introduce other characters to garner support is consistent with the absence of human content in the Rorschach. The helplessness displayed by the character and the narrator is consistent with the positive Coping Deficit Index (CDI) on the Rorschach, suggesting insufficient resources to formulate responses to demands. The short and simplistic story is devoid of inferential or interpretive processes as consistent with vague and concrete information processing on the Rorschach (Lambda, Zd, DQ+, DQv, W:M).

Card 2. The man in the field working. Lady holding onto the stump because she looks like she's pregnant. Girl just come home from school. (T&F). Probably thinking about all the work they gotta do. And the man and the girl probably feeling the same pain, or sorta, for the lady that's pregnant. (Before?) The man started working. (TO?) Everything got done.

The story is closely anchored to the stimulus, yet the narrator does not give priority to the young woman in the foreground, nor does he indicate how the three people are related. Connections are vague and concrete such as the description of the lady as holding on to the stump because she's pregnant. Characters are differentiated only by external appearances, and they all feel the same. As in the previous story, inner life is not elaborated. Carl's Rorschach record contains no human movement, a variable that is associated with greater acceptance of inner thoughts and greater interpersonal awareness. None of the TAT stories suggests the availability of these resources.

Card 3. Little kid was probably tired so he or she fell out the chair. (T/F?) Maybe, I guess they think that's where they gonna sleep at. And probably feeling stupid because he's on the floor. (TO?) That's where they gonna be.

The character falls because he's tired and, despite feeling stupid, just stays there. Again, there's no purposeful action or inner resources but an inertia that's consistent with vague and insufficient processing of information. Feeling stupid, yet doing nothing, is consistent with low self-worth shown on the Rorschach. Other Rorschach variables suggest that self-perceptions are naive and not guided by insight (no FD, no H) nor are these perceptions accompanied by dysphoric mood, anger, or irritation.

Card 4. Lady trying to convince her husband that she love him, but he ain't trying to hear it. (Before?) They was fussing, probably. (TO?) He's about ready to walk away.

The story shows no attempt to understand the inner life or concerns of others but indicates detachment from relationships. The explanation of the stimulus is incongruous—the man looks very angry, but his wife is declaring her love. When asked what happened before, the couple was described as fussing, but the nature of their disagreement is unstated. The picture shows the man turning away, and the narrator does not have the inner resource to provide a solution to the conflict that departs from the stimulus. The Rorschach also suggests that social relationships are superficial, distant, and guarded. The oversimplifying style is likely to lower sensitivity to the needs and interests of others, and this tendency is coupled with a lack of interest or detachment from relationships (no H; no COP or AG). The affect cluster suggests an approach to the environment that minimizes affective engagement. Lack of resources rather than affective provocation is at the heart of Carl's difficulties. His reluctance to process emotional stimuli (Afr) is consistent with his general style of oversimplified processing of information (low blends, high lambda, underincorporating approach).

Card 5. The lady must have heard someone in the house, so she ran and hide in the closet. (T/F?) She probably thinking that she gonna get hurt. (TO?) That no one was there.

The woman runs and hides because she feels vulnerable. Later, when she realizes there's no danger, she makes no attempt to discover the source of her concern (noise). This absence of initiative and curiosity is consistent with nonreflection, simplified processing, and detachment evident in the Rorschach as described previously.

Card 6BM. Looks like his mother told him some bad news. But the news didn't only hurt him; it hurt both of them. (TO?) That something did happen.

Again, the story offers vague, nonspecific descriptions such as bad news and little differentiation between characters. Since the nature of the bad event is not understood, it is not possible to deal with it adaptively. The absence of specific story details is consistent with the other cards and with vague, simplified processing suggested by the Rorschach. The story ends with the character convincing himself that something did happen.

Card 8BM. Looks like they was trying to rob the lady, and she shot one of them. And the man's two buddies trying to get bullets out of him. (T/F?) Probably thinking that he ain't gonna make it. (TO?) That he made it.

The story presents an unrealistic sequence of events and poor integration of foreground and background components of the picture. There's no reasoning about intentions or consequences. Rather, the narrator gives simplistic associations to the stimulus without rule governed connections among likely sequences of events, between causes and effects, or between short and long term outcomes. This detachment from conventional thinking is consistent with poor reality testing on the Rorschach (poor form quality, low populars, incomplete processing of available information).

Card 12M. This is a lady? (It can be whatever it looks like to you) The old man is praying for the lady because he thinks that if he prays, she'll get better. (Before?) She was sick. (TO?) That she didn't make it.

As with the other stories in this protocol, the story does not depart from the immediate cues of the stimulus. Connections between sequences of events and among characters depicted are vague or nonexistent. For example, we do not know what is the relationship between the old man and the lady. The narrator is powerless to introduce alternatives that are not cued by the stimulus such as consulting a physician. The lack of integration and insufficient organization of details is consistent with Rorschach patterns suggesting that Carl formulates decisions without sufficient processing of information (e.g., high lambda, low developmental quality, low blends, and underincorporating style).

Card 13MF. The old man just got up for work, but his wife didn't have breakfast made for him. (T/F?) One of them feeling really tired. The other feeling left out. (TO?) Turns out that one of them sleep and one of them woke.

This story is also rather concrete and stereotypic. Each character has his and her roles to play, but the narrator cannot develop the story beyond what is seen in the stimulus (reliance on immediate external circumstances).

Card 13B. Little boy ain't got no friends and wants something to drink bad. (Before?) Just sitting there. (T/F?) I guess he thinking why he's the only one there. (TO?) That he wasn't.

Just as in Card 5, the ending contradicts the initial premise of the character. Things are not the way they seem, but there's not enough initiative to investigate or deal with the discrepancy. Such lack of investment in processing information is compatible with the failure to process critical cues (underincorporation) on the Rorschach. Again, no purpose, interpersonal connection, or guiding principle is expressed.

The similarity of the conclusions derived from both techniques suggests that they are assessing common processes. Both methods suggest that resources and reality testing are not adequate to meet daily life demands. Furthermore, the focus on the immediate and the concrete, along with haphazard processing of information, hinder the development of schema that are sufficiently elaborated to provide rules that govern the synthesis of experience, the regulation of behavior, and the expression of affect. Without such inner guides to self-regulation, Carl exhibits tendencies toward impulsive, antisocial behavior and has limited capacity to delay gratification. Carl would benefit from a highly structured learning environment with highly structured tasks and frequent feedback and redirection.

4.16.13 VALIDATION ISSUES

Major criticism has been directed not only towards projective methods but also towards psychometric approaches to assessing personality which classify instances of experience, thought, or action into trait categories. The inference process, in general, has been maligned by the behaviorist movement. Compelling arguments have been made to exclude from consideration anything but overt behavior and objectively coded environmental variables (Skinner, 1953). Psychology has moved far from that position to a recognition of the importance of understanding the meanings that individuals assign to environmental events and to their own behaviors. Indeed, cognitive structures and processes are now considered to be at the heart of individual differences in experience, thought, and action, both adaptive and maladaptive (Cantor & Kihlstrom, 1987).

The process of inference is not restricted to the use of projective tests—it is what makes the psychologist a professional rather than a psychometrist. By giving priority to some scores or using qualitative information to increase understanding of the scores, the clinician engages in the process of making professional judgments (Groth-Marnat, 1990). These judgments cannot and should not be eliminated. It is important, however, for practitioners to reflect on the quality of their inferences and the usefulness of their decisions.

Evidence for validity of conclusions made on the basis of inferences drawn from projective methods needs to be convincing. Absolute standards that cut across all methods are essential and they must be precise enough to inform judgments about when the standard is or is not being met. However, in their application to projective techniques, general criteria should accommodate to the nature of the method. Validity and reliability are not established for a generic technique such as figure drawings, the TAT, or the Rorschach. Rather, the utility of each set of instructions and stimuli together with the scoring method is separately evaluated according to the accomplishment of its specific purpose. Ways of demonstrating validity and reliability for all techniques must be in tune with the logic and coherence of the measures and constructs under consideration.

4.16.13.1 Reliability

Establishing reliability for projective techniques and for questionnaire methods is fundamentally different. As Kelly (1958) pointed out: "When the subject is asked to guess what the examiner is thinking, we call it an objective test; when the examiner tries to guess what the subject is thinking, we call it a projective device" (p. 332). Reliability of items on a rating scale is essentially a matter of consistency in the respondent's interpretation of the items. Reliability of inferences based on projective methods rests on a combination of the stimulus, the response, the method of interpretation, and the skill of the interpreter. Psychometric procedures to establish reliability based on forced-choice test items can be carried out without any clinical expertise because reliability pertains primarily to the test-takers responses. With projective measures, reliability, in part, is an attribute of the interpreter because the scoring units are inferences of the professional. Attempts to designate cookbook procedures so that the validation process can be carried out quickly and easily are out of tune with the nature of projective methods.

The following is a brief overview of how traditional psychometric indicators must be modified to establish reliability of projective techniques:

4.16.13.1.1 Scorer reliability

The accuracy of two people looking for the same information requires adequate training and clear guidelines. When raters are well-trained and scoring systems are well documented, such reliabilities tend to be high. A related aspect of scorer reliability, particularly relevant to clinical inference is the consistency with which one rater codes the same protocol over time (Karon, 1981).

4.16.13.1.2 Decisional reliability

The consistency of decisions drawn from a protocol can be assessed when specific units are not the focus. For example, Shneidman (1951) showed that 16 clinicians using their own methods came to similar conclusions. Other influences on the reliability of decisions may relate to the number of performance samples needed. For example, in the measurement of particular motives with TAT cards, it is important to know if the specific motive emerges reliably in every card or only some. To assure adequate reliability, it has been recommended that at least six stories be obtained from each respondent (Lundy, 1985; Smith, 1992b).

Decisions must be based on clear conceptualizations. Therefore, decisional reliability must account for the theoretical appropriateness of the match between predictor and criteria. Test responses can be expected to correlate with criteria only if their meaning is functionally similar. Thus, adequate functioning in a structured situation can occur despite disorganized responses on a projective test.

4.16.13.1.3 Test–retest reliability

One important factor in demonstrating the reliability of the measure upon retesting is whether one is looking for similarity of content or of the inference (Karon, 1981). The reliability of the specific content is less relevant than consistency in the meaning of the response.

4.16.13.1.4 Internal consistency

Internal consistency of thematic content across cards would be an inappropriate measure of reliability because cards are designed to elicit different themes (Lundy, 1985). However, despite variability in specific content, the responses may yield similar inferences. More stylistic units representing psychological processes such as the accuracy with which the content captures the "gist" of the scene presented or linkages between causes and effects may be generalizable across different stimuli. Internal consistency based on the number of words per story (alpha = 0.96) was much higher than internal consistency of need for achievement (Atkinson, Bongort, & Price, 1977). Likewise, Rorschach cards present stimuli with important differences and, rather than estimating internal consistency, inferences are made on the basis of the entire coded protocol and not card by card. Responses to the white spaces in the blot have different meaning for each card, and not all white space responses represent figure ground reversal. Therefore, adding all space responses is a rough estimate of the trait in question.

Internal consistency expected within a measure should relate to the nature of the task and to the consistency inherent in the construct under consideration.

4.16.13.2 Construct Validation

Messick (1989) defines construct validity as "an integration of any evidence that bears on the interpretation or meaning of test scores" (p. 17). Because traditional indices of content or criterion validity contribute to the meaning of test scores, they too pertain to construct validity. Thus, construct validity subsumes all other forms of validity evidence.

This emphasis on construct validity as the overriding focus in test validation represents a shift from prediction to explanation as the fundamental focus of validation efforts. Construct validation emphasizes the development of models explaining processes underlying performance on various tests and their relationships to other phenomena. Accordingly, correlations between test scores and criterion measures contribute to the construct validity of both predictor and criterion. In Messick's words, "Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment" (p. 13).

4.16.13.3 Multitrait, Multimethod Validation

Although the multitrait, multimethod approach is a standard technique for construct validation (Campbell, 1960; Campbell & Fiske, 1959), this technique is vulnerable to problems with the definition and measurement of constructs. This procedure seeks to establish higher correlations across diverse measures of the same trait (convergent evidence) and lower correlations among similar measures of different traits (discriminant evidence) to show that a construct is distinct from other constructs and that it is not uniquely tied to a particular measurement method. Yet, the possibility that the constructs may indeed be tied to the measurement tool must be considered. McClelland and colleagues showed that attempts to correlate self-report and projective measures of achievement motivation to validate either one are misguided. Likewise, Horowitz (1991) notes that schema, although important to measure, are not directly known by the subject and cannot be confirmed through self-report but by a consensus of two or more independent observers. Theoretical understanding of the construct must always be central to the choice of validation efforts.

It is important to acknowledge the potential for corroboration of constructs with various assessment measures. Yet, different measures of the same construct (e.g., ratings by actor versus observer) may be assessing qualitatively different aspects of that construct. Meaningful differences in the constructs are demonstrated by distinct patterns of relationships with various criterion measures (see criterion validation below).

Even highly similar measures (e.g., thematic approaches) used to assess identical constructs but with different procedures are not necessarily addressing the same qualities. For example, Arnold (1962) developed criteria for scoring TAT stories for achievement motivation by comparing groups of individuals known to differ in their job success. In contrast, the McClelland–Atkinson tradition (McClelland & Koestner, 1992) for developing criteria for scoring achievement motivation from stories told to picture stimuli was based on the comparison of groups given different instructional sets to arouse the achievement motive. Each of these two approaches to contrasting groups may be appropriate for given purposes with specific populations but clearly involve distinct conceptualizations of the achievement motive. One takes the position that the motive is present and, when aroused, energizes and directs behavior towards a particular class of goals or incentives. The other focuses on the achievement motive as being a relatively stable disposition reflected by complex patterns of cognitive–emotional processes that guide perceptions and behavior. If scoring units are empirically derived through contrasting groups, then the construct is defined by the nature of the groups and procedures used.

4.16.13.4 Criterion Validation

Criterion related evidence also belongs under the rubric of construct validity (Messick, 1989). Patterns of correlations with other variables

contribute to the defining attributes of each construct. The same set of external correlates do not apply to achievement motivation measured through projective and self-report methods (Spangler, 1992). Projective measures of achievement motivation correlate with spontaneous effort sustained over time, whereas self-reports correlate with activities that are cued by the situation. Such variation in the match between the predictor and the target clarify the dimensions of the achievement motive assessed by each type of measure. This understanding of constructs in terms of the life conditions to which they generalize constitutes the simultaneous validation of the predictor and the criteria. Relationships between predictors and criteria derived from a construct must be tested for expected patterns of discriminant and convergent evidence. In doing so, both situational specificity and generalizability of constructs and their measures are addressed at the same time. Generalizability of the predictor–criterion relationships across different population groups, settings, task domains, ages, or gender must be empirically established.

4.16.13.5 Part–Whole Relationships

Personality is a functional whole comprised of patterns of interrelated components. Therefore, predictions from single elements of personality to more encompassing units, such as general adjustment, proceed from part to whole and can only account for a portion of the variance. This part–whole prediction is illustrated in Figure 1 showing the influence of temperament on adjustment as mediated by goodness-of-fit. Any one temperamental attribute would account for a small proportion of the variance in the fit between the person and various situational contexts or task demands. A configuration of such traits would improve the prediction. However, the actual goodness-of-fit would be a better predictor of adjustment because this broader construct subsumes the configuration of traits (even those not specifically addressed in the prediction), the environmental demands and supports, as well as coping mechanisms.

This understanding of the part–whole relationship is essential for the validation of projective techniques. Projective techniques look at individual differences at a more global level than other measures. The products reflect schema or inner structures that develop through the individual's synthesis of life experiences. These involve the interplay of all of the person's characteristics and reciprocal transactions with the environment. Therefore, part–whole patterns exist between various trait constructs and response parameters within the test. Part–whole predictions can relate specific neuropsychological or temperamental processes to individual differences in responses to projective tasks (Bassan-Diamond, Teglasi, & Schmitt, 1995). Part–whole relationships depend on conceptualizations about how the various dimensions of individual differences come together in the development of broader units of personality such as the influence of temperament on conscience development (Kochanska, 1993). Given the expected role of multiple variables, it is apparent that expectations about the size of correlations need to be in line with understanding of the phenomena.

The part–whole issue also applies to the relationships among responses within a specific projective technique. The functional significance of each part hinges on its interplay with other parts in relation to the constructs being considered. Various dimensions of the product such as form and content variables are cohesively related and need to be systematically coordinated into higher-order constructs. Psychometric and theory-based approaches need to be integrated to build useful frameworks for clinical use. Specific response parameters can be abstracted from the whole and pieced together to find conceptually meaningful patterns.

4.16.13.6 Convergence of Psychometric and Conceptual Treatment of Data

Normative, nomothetic, and case study approaches to data collection are interrelated. Conceptual linkages among these approaches are essential to the interpretation and validation of projective techniques, as described next.

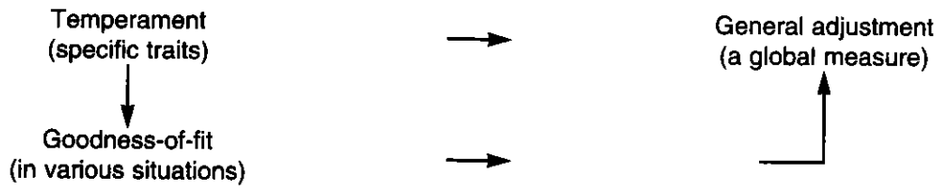


Figure 1 Illustration of part–whole prediction.

4.16.13.6.1 Normative

Before initiating the arduous process of collecting norms for projective methods, the units of analysis must be conceptually meaningful, theoretically relevant, and clinically useful. Furthermore, criteria for coding must be provided in sufficient detail to assure rater reliability. The Comprehensive System for the Rorschach has shown that the establishment of norms for meaningful units of analysis is possible, and that larger conceptual units can be designated through the use of multiple cutoffs to establish interpretively useful patterns of interrelated response parameters.

4.16.13.6.2 Nomothetic

It is important to base norms on units of interpretation that represent psychological processes that can serve as explanatory constructs. The nomothetic approach attempts to promote conceptual understanding through the study of general principles and functional relationships among variables. Such a conceptual approach to the Rorschach (Blatt & Berman, 1984; Weiner, 1977) has advocated organizing discrete response parameters into theoretically relevant clusters that contribute to a well defined construct. A strict empiricist would be satisfied with the prediction of patterns of behavior from various configurations of test responses (e.g., of Rorschach variables). However, establishing such empirical relationships is only a starting point for building conceptual frameworks for understanding and explaining behavior. The meaning of test response patterns emerges from a network of relationships with other theoretically relevant response patterns and theoretically appropriate external criteria.

The collection of normative data for projective techniques is embedded in the construct validation process. The constructs and units of inference for projective measures need to be defined *a priori* in the same way that test items in questionnaires are selected to represent constructs. An example of this approach is the designation of units of inference for TAT stories on the basis of constructs derived from research and theory on empathy (Locraft & Teglasi, 1997). Scores based on these units subsequently differentiated groups of children designated as high, medium, and low on empathy on the basis of teacher ratings. Such units require cross validation on numerous populations to assure wide applicability prior to attempting large scale normative studies. Unless the unit of inference is conceptually clear and represents meaningful psychological processes, normative data or group comparisons are not particularly useful.

4.16.13.6.3 Case study

The case study is most relevant to the practicing clinician because decisions for clinical purposes relate to understandings about one person. The adequacy of these decisions can be validated on a case by case basis. Although case study methods focus on the unique patterns of variation in responses of a single individual, the pattern of variables under consideration can be referenced to norms and to theoretical constructs. With projective methods, the respondent is allowed free expression within a specified context (stimuli and directions), and effective use of these techniques require a conceptual framework at three levels. First is an understanding of the psychological meaning of the response patterns within the projective measure. Second is an understanding of how the projective measure fits with other information in a comprehensive evaluation. Third is a conceptualization of how the patterns relate to competencies needed in the relevant life situations. Therefore, to conduct adequate case studies, the examiner needs a framework to understand how various types of data fit together within and across measures and how the emerging patterns apply to the various demands of the individual's environment. The attempt to integrate psychometric (empirical) and nomothetic (theoretical) data into the case study is typified in the attempt to synthesize structural and content features of Rorschach responses (Erdberg, 1993). The problem is that these efforts have applied different conceptual frameworks to different aspects of the data. A true integration of the empirical and theoretical perspectives requires a focus on personality constructs and the establishment of coherent patterns of data from various sources around the constructs.

The term "conceptual validity" has been applied to the process of psychological assessment (Maloney & Ward, 1976). Whereas construct validity focuses on confirming expected patterns of relationships across individuals, conceptual validity focuses on observing cohesive patterns within an individual. These patterns of expected relationships among observations constitute a working model of the individual being evaluated. According to Maloney and Ward, establishing such a model is a prerequisite for answering the referral question. When information from various sources is understood in terms of the constructs that explain an individual's difficulties and point to appropriate decisions, the assessment has conceptual validity.

4.16.14 FUTURE DIRECTIONS

The anticipation of future directions for projective techniques emerges from the clues gleaned from the scientific literature pertaining to the development and assessment of personality. However, the prognosticator's wishful thinking also influences this process. The intent of the prognosticator is to show how unfolding trends or even new spins on old ideas point toward desired alternatives for moving forward. Future possibilities for projective techniques are drawn from three conceptually distinct perspectives: (i) converging constructs from the various subfields of psychology that scaffold the use of projective techniques; (ii) refinements in the conceptualization of personality that point to a multidimensional view and to a unique role for projective techniques; and (iii) improved understanding and use of the specific projective tools.

Methods and constructs in other subfields of psychology provide a hospitable *zeitgeist* for projective techniques and bode well for their future development and utility. Converging evidence appears to validate the basic tenets of the projective hypothesis. Concepts from psychodynamic formulations such as transference have been redefined in terms of contemporary psychology, demonstrating that such phenomena are not unique to one theory but can be understood in several ways (Singer & Singer, 1994). The examination of clinically relevant phenomena in light of conceptualizations of memories, schema, and scripts from various subdisciplines support the work in each subfield (e.g., Horowitz, 1991; Stein & Young, 1992). Cognitive theories of perception, memory, and learning are increasingly emphasizing the unconscious or implicit social attitudes (Uleman & Bargh, 1989). Research on construct availability and accessibility (Higgins, 1990) as well as script or schema theory also recognizes the influence of cognitive processes that occur outside of awareness. Knowledge from other areas of psychology will increasingly inform the designation of interpretive units for measuring personality with projective techniques. Recent interest in narrative methods to assess schema and social cognitions is applicable to thematic apperceptive techniques (Cramer, 1996). Stories elicited through picture stimuli, like other narrative approaches, reflect the human orientation to perceive the world as stories or myths grounded in culture and personal experience.

Conceptions of psychopathology within psychoanalytic theory are placing increasing emphasis on functions of inner psychic structures and subjective meaning systems (Atwood & Stolorow, 1984). In this context, adjustment problems arise because the structures are insufficiently developed to permit adaptive coping rather than from conflict among the structures. In an analogous manner, clinical syndromes have been related to problems in the development of person schema (Horowitz, 1991). If psychopathology is a reflection of impairment in the formation of psychic structures, then it is reasonable for assessment to be geared to the evaluation of these structures (e.g., object relations, self-system, along with associated processes such as reality testing).

Levels of impairment based on the complexity and organization of inner structures will be identified along with a description of symptomatic behaviors. Projective techniques will figure prominently in the assessment of psychological variables for the purpose of intervention planning. The continued refinement of integrative therapies (Leve, 1995; Norcross, 1986) will increase the usefulness of identifying relevant processes to be targeted and matched with optimal intervention strategies.

Understanding personality according to different styles and levels of organization of experience as they relate to functioning will rest on increasing emphasis on part whole conceptualizations. For example, with a focus on inner structures, aspects of personality such as emotions will be increasingly recognized as organizing processes that shape adaptation and problem solving (Greenberg, Rice, & Elliott, 1993). Other personality processes, such as distractibility or inattention, will also be understood in relation to how they shape the development of inner psychic structures. Future research will focus on establishing linkages between various levels of personality constructs such as the role of various traits in the development of inner strivings and the organization of meaning structures. Research will also focus on clarifying relationships between discrete measures of neuropsychological processes, such as planning, organizing, retrieval from memory, or continuous attention, and the manifestation of these processes in the performance of more complex tasks analogous to real-life situations. Inherent in the part whole conceptualization of personality is the view of development as being propelled simultaneously by biological, psychological, and social forces. Within the psychological realm, perceptions and behaviors will be increasingly understood in terms of the interplay of experience with relevant affective and cognitive processes.

A multilevel understanding of personality will permit clearer conceptualizations of the utility of different types of techniques to assess different facets of personality. One example is the distinction between measures of self-attributed and implicit motives to achieve (McClelland et al., 1989). Story-telling measures of achievement motivation (implicit) were more effective in the prediction of long term outcomes such as career success, whereas self-reports (self-attributed) were better in predicting immediate choices (Spangler, 1992). The two types of measures of achievement motivation relate to different criteria, develop through different pathways, and can be understood in terms of different levels of personality.

In training programs, the professional courses are the places where the core psychological knowledge bases are integrated. Expertise in projective techniques includes the linking of the psychological processes assessed with appropriate strategies for therapeutic interventions. The complexity of interpreting projective instruments and of understanding their implications for guiding interventions requires no less than the systematic and flexible application of prior knowledge. Therefore, projective assessment can serve as a centerpiece for integrating concepts from core areas such as physiological, affective and cognitive bases of behavior. Training should emphasize the gradual development of the professional's schema to guide practice rather than the acquisition of knowledge applied in rote fashion. Trainees' schema should be sufficiently broad and flexible not only to incorporate information currently available but to accommodate continuous learning throughout the professional career. Measures are simply tools. Rather than emphasizing the assessment technique, training programs would do well to focus on the development of integrative frameworks that include how various sources of information fit together into meaningful patterns.

Increased acceptance of qualitative measures as having scientific merit has led to more frequent use of open-ended methods in the study of personality such as the study of early memories (Bruhn, 1992), thought sampling (Rubin, 1986), examination of the life story (McAdams, 1990), self-defining memories (Moffit & Singer, 1994), and analysis of therapy transcripts (Luborsky & Crits-Christoph, 1990). The psychometric challenges of projective methods are shared among all open-ended techniques. The effort to master these difficulties will require, above all, conceptual clarity and an emphasis on construct validation. These endeavors will promote shared conceptualizations and methods across disciplines, appreciation of the progress already made with projective techniques, and the spurring of further developments.

Frameworks for interpretation of projective techniques not only elucidate the inner structure of personality but reveal problem-solving strategies in relation to the task demands. Therefore, it may be fruitful to view these techniques, in part, as *performance measures of personality* that can be compared to other tests in an assessment battery in terms of what the task requires. The most significant difference is that performance tasks of personality maximize the use of spontaneous strategies to organize perceptions and responses. Variation in response patterns across different types of tasks permits distinctions between knowledge structures that are inert and require external prompts from those that are meaningfully organized and spontaneously accessible. Furthermore, the configuration of responses within and across diverse tasks foster explanations of situational variability and consistency in performance and behavior. For instance, social situations require the individual to organize perceptions, size up intentions, and anticipate reactions of others. These requirements are similar to the demands of projective tests to organize perceptions of the stimuli and coordinate the responses with the instructions.

The relationship between content and formal elements in the units of inference drawn from projective techniques still needs to be clarified. The most fruitful approach to integrating the analysis of form and content is through the identification of their linkages with important psychological processes. Formal structural analysis should receive greater emphasis in the apperceptive methods. Generalized properties of content and features of narrative structure are more durable indices of psychological processes than isolated content. Furthermore, the analysis of the subtext or underlying structure of narratives can yield significant information about the individual's schematic organization of experience. The struggle to apply valid and reliable methods to assessment will continue. The inevitable link between theory and method necessitates the simultaneous effort to refine both. Conceptualizations developed in other subfields of psychology and lessons learned from qualitative research methods will be applied to improve the reliability, validity, and clinical utility of projective measures.

4.16.15 SUMMARY

The chapter is summarized by presenting a model for understanding performance measures of personality shown in Figure 2. The interaction of the person and environment (boxes 1 and 2) as a fundamental unit of study has wide acceptance and applicability. However, the objective features of the environment and the subjective

world are not the same. Lewin (1935) refers to the psychological environment as the inner experience of external reality. Persons participate in and influence their external environments as well as being shaped by them. The manner in which the individual encodes and stores encounters with the external world drives the construction of the inner psychological world (box 3). Therefore, one can only conceptually separate the person from the environment—they are embedded in one another.

Schema are memory structures that develop through the individual's constant and reciprocal transactions with the environment. The development of schema is influenced by (all factors said to influence personality) various interactive determinants including constitutional factors of temperament (see reviews by Emde, 1989; Plomin, 1986) as well as meaning structures transmitted by family and culture in conjunction with the individual's experiences throughout life. These schema are important to assess because they: (i) are mental sets that guide attentional processes and serve as filters for interpreting information; (ii) guide actions in situations that require complex resources; and (iii) shape the storage of new information in memory.

Certain conditions such as attention-deficit hyperactivity disorder are diagnosed through a careful examination of previous history as well as scrutiny of current behavioral patterns. However, these attentional processes shape the individual's prior and ongoing synthesis of experience through their impact on the information entering awareness, the feedback received from others and on the development of internalized schema. Therefore, the part–whole relationships between specific symptoms or psychological processes and larger units of the personality need to be understood. Projective testing relies on an appreciation that overt behavior is linked to inner meaning. Consequently, inner life is a mechanism for perceiving the outer world and for regulating behavior to adapt to these perceptions. The projective techniques (e.g., TAT and Rorschach) provide information about how prior experience has been organized and how inner resources (schema, inner structures) are applied to relatively unstructured situations. Responses to projective tasks are understood as cohesive products reflecting all of the processes involved in the synthesis of experience. Specific psychological processes such as attention or emotion can be studied as precursors and sequelae of the development of inner structures.

Projective techniques require a person to demonstrate the qualities of organization and strategic planning by performing a problem-solving task rather than by telling about the self in an interview or by responding to a questionnaire. Therefore, the term *performance measure of personality* aptly describes projective methods. These performance measures are less structured than the other tasks in the typical assessment battery and provide unique information. The manner in which the problem is solved or the task is accomplished reveals the individual's organizational strategies and resources to deal with similarly unstructured situations. In addition, the nature and organization of knowledge structures reveal the frameworks or convictions that guide responses in ambiguous situations. Furthermore, inferences drawn from performance measures and self-reports pertain to different facets of personality and predict different criteria (McClelland et al., 1989).

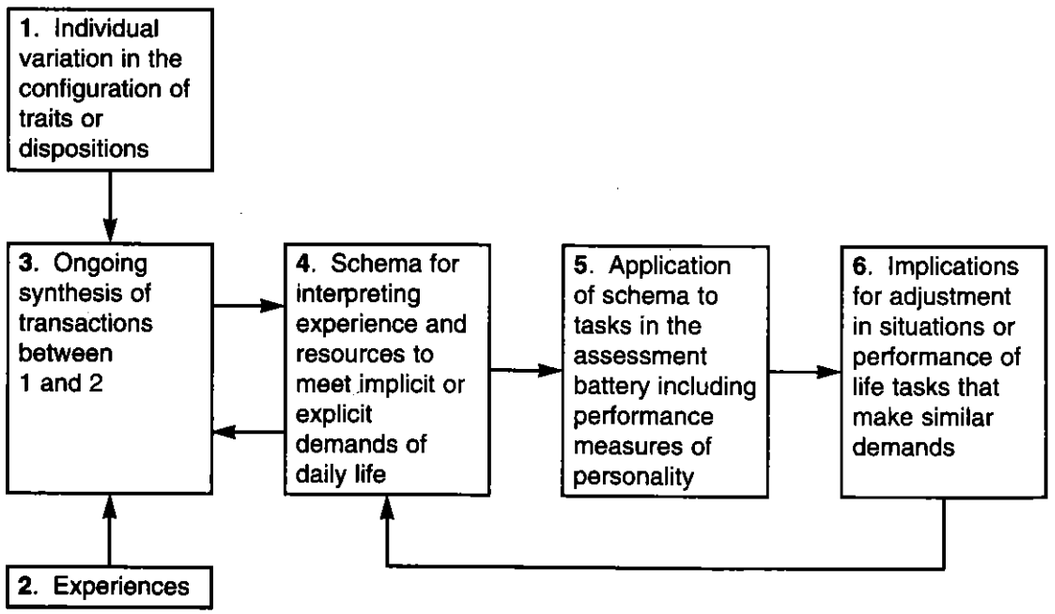


Figure 2 Understanding performance measures of personality.

As Figure 2 shows (box 4), the individual's ongoing synthesis of experience leads to the development of knowledge structures. Various performance measures in an assessment battery can be examined in terms of what the task requires and what inner structures or resources are brought to bear on the performance (box 5). Actual adjustment in a given situation (box 6) depends not only on the person's resources but also hinges on environmental expectations and supports.

The person is assessed within a particular social context of the testing situation including expectations about testing, anticipation of the potential consequences of testing, and the relationship with the examiner. Therefore, these variables are considered along with the task demands. The job of the examiner is to organize multiple sources of data including reports of self and others, various performance measures, relevant past history, and current circumstances into a framework for understanding the referral issues. All of the information is sifted through an understanding of family and cultural background as pertains to the individual being assessed. The understanding is informed not only by theories of personality, human development, and psychopathology, but also by a conceptualization of the various tasks in the battery and how they relate to performance in life situations. Findings are not a listing of isolated facts but conclusions based on cohesive patterns that provide an understanding of the problem and point to appropriate action.

4.16.16 REFERENCES

Abelson, R. P. (1976). Script processing in attitude formation and decision making. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior*. Hillsdale, NJ: Erlbaum.

Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist, 36,* 715–729.

Acklin, M. W. (1994). Some contributions of cognitive science to the Rorschach Test. In I. B. Weiner (Ed.), *Rorschachianna* (pp. 129–145). Seattle, WA: Hogrefe & Huber.

Alper, T. G., & Greenberger, E. (1967). Relationship of picture structure to achievement motivation of college women. *Journal of Personality and Social Psychology, 7,* 362–371.

Anastasi, A. (1976). *Psychological testing*. New York: Macmillan.

Anderson, J. R. (1990). *Cognitive psychology and its implications*. New York: Freeman.

Archer, R. P., Maruish, M., Imhof, E. A., & Piotrowski, C. (1991). Psychological test usage with adolescent clients: 1990 survey findings. *Professional Psychology: Research and Practice, 22,* 247–252.

Arnold, M. B. (1962). *Story sequence analysis: A new method of measuring motivation and predicting achievement*. New York: Columbia University Press.

Atkinson, J. W. (1992). Motivational determinants of thematic apperception. In C. P. Smith (Ed.), *Motivation and personality: Handbook of thematic content analysis* (pp. 21–48). New York: Cambridge University Press.

Atkinson, J. W., Bongort, K., & Price, L. H. (1977). Explorations using computer simulation to comprehend TAT measurement of motivation. *Motivation and Emotion, 1,* 1–27.

Atwood, G., & Stolorow, R. (1984). *Structures of subjectivity*. Hillsdale, NJ: Analytic Press.

Auld, F., Jr. (1954). Contributions of behavior theory to projective testing. *Journal of Projective Techniques, 18,* 421–426.

Bargh, J. A. (1989). Conditioned automaticity: Varieties of automatic influence on social perception and cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 3–51). New York: Guilford Press.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 1–40). Hillsdale, NJ: Erlbaum.

Barkley, R. A. (1990). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment*. New York: Guilford Press.

Bassan-Diamond, L. E., Teglasi, H., & Schmitt, P. (1995). Temperament and a story-telling measure of self-regulation. *Journal of Research in Personality, 29,* 109–120.

Baumeister, R. F., Wotman, S. R., & Stillwell, A. M. (1993). Unrequited love: On heartbreak, anger, guilt, scriptness, and humiliation. *Journal of Personality and Social Psychology, 64,* 377–394.

Beck, S. J. (1981). Reality Rorschach and perceptual theory. In A. I. Rabin (Ed.), *Assessment with projective techniques* (pp. 23–46). New York: Springer.

Bellak, L. (1975). *The T.A.T., C.A.T., and S.A.T. in clinical use*. New York: Grune & Stratton.

Bellak, L. (1993). *The T.A.T., C.A.T., and S.A.T. in clinical use* (5th ed.). Needham Heights, MA: Allyn & Bacon.

Birney, R. C. (1958). Thematic content and the cue characteristics of pictures. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society* (pp. 630–643). New York: Van Nostrand.

Blatt, S. J. (1991). A cognitive morphology of psychopathology. *Journal of Nervous and Mental Disease, 179,* 449–458.

Blatt, S. J. (1992). The Rorschach: A test of perception on an evaluation of representation. In E. I. Megargee & C. D. Spielberger (Eds.), *Personality assessment in America: A retrospective on the occasion of the fiftieth anniversary of the Society for Personality Assessment* (pp. 160–169). Hillsdale, NJ: Erlbaum. (Reprinted from 1990 *Journal of Personality Assessment, 55,* 394–416).

Blatt, S. J., & Berman, W. H., Jr. (1984). A methodology for the use of the Rorschach in clinical research. *Journal of Personality Assessment, 48,* 226–239.

Blatt, S. J., Ford, R. Q., Berman, W., et al. (1988). The assessment of therapeutic change in schizophrenic and borderline young adults. *Psychoanalytic Psychology, 5,* 127–158.

Blatt, S. J., & Lerner, H. (1983). The psychological

assessment of object representations. *Journal of Personality Assessment, 47,* 7–28.
Blatt, S. J., & Wild, C. M. (1976). *Schizophrenia: A developmental analysis.* New York: Academic Press.
Blatt, S. J., Allison, J., & Feirstein, A. (1969). The capacity to cope with cognitive complexity. *Journal of Personality, 37,* 269–288.
Bransford, J. D., Franks, J.J., Vye, N. J., & Sherwood, R. D. (1989). New approaches to instruction: Because wisdom can't be taught. In S. Voshiadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 470–497). Cambridge, UK: Cambridge University Press.
Bruhn, A. R. (1992). The early memories procedure: A projective test of autobiographical memory, Part 2. *Journal of Personality Assessment, 58,* 326–346.
Bruner, J. S. (1986). *Actual minds, possible worlds.* Cambridge, MA: Harvard University Press.
Buck, J. N. (1948). The H–T–P technique: A quantitative and qualitative scoring manual. *Clinical Psychological Monographs, 5,* 1–120.
Buck, J. N. (1987). *The House-Tree-Person technique: Revised manual.* Los Angeles: Western Psychological Services.
Burns, R. C. (1987). *Kinetic-House-Tree-Person drawings (KHTP).* New York: Brunner/Mazel.
Burns, R., & Kaufman, S. (1970). *Kinetic Family Drawings (K-F-D): An introduction to understanding children through kinetic drawings.* New York: Brunnel/Mazel.
Burns, R. C., & Kaufman, H. S. (1972). *Actions, styles, and symbols in Kinetic Family Drawings: An interpretive manual.* New York: Brunner/Mazel.
Butcher, J. N., & Rouse, S. V. (1996). Personality: Individual differences and clinical assessment. In J. T. Spence, J. M. Darley, & D. J. Foss (Eds.), *Annual Review of Psychology, 47,* 87–111.
Campbell, D. T. (1960). Recommendations for APA test standards regarding construct, trait or discriminant validity. *American Psychologist, 15,* 546–553.
Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105.
Campus, N. (1976). A measure of needs to assess the stimulus characteristics of TAT cards. *Journal of Personality Assessment, 40,* 248–258.
Cantor, N., & Kihistrom, J. F. (1987). *Personality and social intelligence.* Englewood Cliffs, NJ: Prentice-Hall.
Carlson, L., & Carlson, R. (1984). Affect and psychological magnification. Deviations from Tomkins' script theory. *Journal of Personality, 52,* 36–45.
Chi, M., Glaser, R., & Farr, M. (1987). *Nature of expertise.* Hillsdale, NJ: Erlbaum.
Clore, G. L., Schwarz, N., & Conway, M. (1994). Affective causes and consequences of social information processing. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (pp. 323–417). Hillsdale, NJ: Erlbaum.
Cooper, A. (1981). A basic TAT set for adolescent males. *Journal of Clinical Psychology, 37,* 411–414.
Cramer, P. (1996). *Storytelling narrative and the Thematic Apperception Test.* New York: Guilford Press.
Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social-information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115,* 74–101.
Cronbach, L. J. (1970). *Essentials of psychological testing* (3rd ed.). New York: Harper & Row.
Demorest, A. P., & Alexander, I. E. (1992). Affective scripts as organizers of personal experience. *Journal of Personality, 60,* 645–663.
Dodge, K. A., & Feldman, E. (1990). Issues in social cognition and sociometric status. In S. R. Asher & J. D. Coie (Eds.), *Peer rejection in childhood* (pp. 119–155). New York: Cambridge University Press.
Durand, V. M., Blanchard, E. B., & Mindell, J. A. (1988). Training in projective testing: Survey of clinical training directors and internship directors. *Professional Psychology: Research and Practice, 19,* 236–238.
Emde, R. N. (1989). The infant's relationship experience: Developmental and affective aspects. In A. J. Sameroff & R. N. Emde (Eds.), *Relationship disturbances in early childhood: A developmental approach.* New York: Basic Books.
Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist, 49,* 709–724.
Erdberg, P. (1993). The U.S. Rorschach scene: Integration and elaboration. In I. B. Weiner (Ed.), *Rorschachianna XIX: Yearbook of the International Rorschach Society* (pp. 139–151). Seattle, WA: Hografe & Huber.
Exner, J. E. (1989). Searching for projection in the Rorschach. *Journal of Personality Assessment, 53,* 520–536.
Exner, J. E. (1993). *The Rorschach: A comprehensive system: Vol. 1. Basic processes.* New York: Wiley.
Exner, J. E., & Weiner, I. B. (1995). *The Rorschach: A comprehensive system: Vol. 3. Assessment of children and adolescents.* New York: Wiley.
Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50,* 229–238.
Feher, E., VandeCreek, L., & Teglasi, H. (1983). The problem of art quality in the use of the Human Figure Drawing Test. *Journal of Clinical Psychology, 39,* 268–275.
Festinger, L. (1957). *A theory of cognitive dissonance.* Evanston, IL: Row, Peterson.
Fisher, S. (1986). *Development and structure of the body image.* Hillsdale, NJ: Erlbaum.
Fiske, A. P., Haslam, N., & Fiske, S. T. (1991). Confusing one person with another: What errors reveal about the elementary forms of social relations. *Journal of Personality and Social Psychology, 60,* 656–674.
Fiske, S. T., & Taylor, S. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.
Frank, L. D. (1939). Projective methods for the study of personality. *Journal of Psychology, 8,* 389–413.
Frank, L. D. (1948). *Projective Methods.* Springfield, IL: Thomas.
Fromkin, H. L., & Streufert, S. (1976). Laboratory experiments. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 415–465). Chicago: Rand McNally.
Goodenough, F. L. (1926). *Measurement of intelligence by drawings.* New York: World Books.
Greenberg, L. S., Rice, L. N., & Elliott, R. (1993). *Facilitating emotional change: The moment by moment process.* New York: Guilford Press.
Groth-Marnat, G. (1990). *Handbook of psychological assessment* (2nd ed.). New York: Wiley.
Hammer, E. F. (1958). *The clinical application of projective drawings.* Springfield, IL: Charles C. Thomas.
Handler, L. (1985). The clinical use of the Draw-A-Person Test (DAP). In C. S. Newmark (Ed.), *Major psychological assessment instruments.* Newton, MA: Allyn & Bacon.
Handler, L., & Habernicht, D. (1994). The Kinetic Family Drawing Technique: A review of the literature. *Journal of Personality Assessment, 62,* 440–464.
Harris, D. B. (1963). *Children's drawings as measures of intellectual maturity.* New York: Harcourt Brace Jovanovich.
Hartman, A. A. (1970). A basic TAT set. *Journal of Projective Techniques, 34,* 391–396.
Haynes, J. P., & Peltier, J. (1985). Patterns of practice with TAT in juvenile forensic settings. *Journal of Personality Assessment, 49,* 26–29.

Heider, F. (1958). *The psychology of interpersonal relations.* New York: Wiley.
Henry, W. E. (1956). *The analysis of fantasy: The Thematic Apperception Test in the study of personality.* New York: Wiley.
Higgins, E. T. (1990). Personality, social psychology, and person-situation relations: Standards and knowledge activism as a common language. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research,* (pp. 301–338). New York: Guilford Press.
Hogan, R. (1987). Personality psychology: Back to basics. In J. Aronoff, A. J. Rabin, & R. A. Zucker (Eds.), *The emergence of personality* (pp. 79–105). New York: Springer.
Holt, R. R. (1958). Formal aspects of the TAT—A neglected resource. *Journal of Projective Techniques, 22,* 163–172.
Holt, R. R. (1961). The nature of TAT stories as cognitive products: A psychoanalytic approach. In J. Kagan & G. Lesser (Eds.), *Contemporary issues in thematic apperceptive methods* (pp. 3–40). Springfield, IL: Charles C. Thomas.
Holt, R. R. (1978). *Methods in clinical psychology: Vol. 1. Projective assessment.* New York: Plenum.
Holt, R., & Luborsky, L. (1958). *Personality patterns of psychiatrists: A study of methods for selecting residents.* New York: Basic Books.
Holtzman, W. H. (1981). Holtzman Inkblot Technique. In A. I. Rabin (Ed.), *Assessment with projective techniques* (pp. 47–83). New York: Springer.
Holtzman, W. H., Thorpe, J., Swartz, J., & Heron, E. (1961). *Inkblot perception and personality: Holtzman Inkblot Technique.* Austin, TX: University of Texas Press.
Horowitz, M. J. (1991). *Person schema and maladaptive interpersonal patterns.* Chicago: University of Chicago Press.
Ingram, R. E. (Ed.) (1986). *Information processing approaches to clinical psychology.* New York: Academic Press.
Isen, A. M. (1984). Toward understanding the role of affect in cognition. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 3, pp. 179–236). Hillsdale, NJ: Erlbaum.
Isen, A. M. (1987). Positive affect, cognitive processes, and social behavior. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 203–253). New York: Academic Press.
Johnston, M. H., & Holzman, P. S. (1979). *Assessing schizophrenic thinking.* San Francisco: Jossey-Bass.
Kahana, B. (1978). The use of projective techniques in personality assessment of the aged. In M. Storandt, I. Siegler, & M. Elias (Eds.), *The clinical psychology of aging.* New York: Plenum.
Kahill, S. (1984). Human figure drawings in adults: An update of the empirical evidence, 1962–1982. *Canadian Psychology, 25,* 269–292.
Karon, B. P. (1981). The Thematic Apperception Test (TAT). In A. I. Rabin (Ed.), *Assessment with projective techniques: A concise introduction* (pp. 85–120). New York: Springer.
Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivations* (Vol. 15). Lincoln, NE: University of Nebraska Press.
Kelly, G. A. (1958). The theory and technique of assessment. In P. R. Farnsworth & Q. McNemar (Eds.), *Annual Review of Psychology, 9,* 323–352. Palo Alto: Annual Reviews.
Kenny, D. T. (1964). Stimulus functions in projective techniques. In B. A. Maher (Ed.), *Progress in experimental personality research* (pp. 285–354). New York: Academic Press.
Kenny, D. T., & Bijou, S. W. (1953). Ambiguity of pictures and extent of personality factors in fantasy responses. *Journal of Consulting Psychology, 17,* 283–288.
Kihlstrom, J. F. (1984). Conscious, subconscious and preconscious: A cognitive perspective. In K. S. Bower & D. Michenbaum (Eds.), *The unconscious reconsidered.* New York: Wiley.
Kihlstrom, J. F. (1987). The cognitive unconscious. *Science, 237,* 1445–1452.
Kihlstrom, J. F. (1990). The psychological unconscious. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 445–464). New York: Guilford Press.
Kihlstrom, J. F., & Cantor, N. (1984). Mental representations of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 17) New York: Academic Press.
Klopfer, W. G. (1981). Integration of projective techniques in the clinical case study. In A. I. Rabin (Ed.), *Assessment with projective techniques* (pp. 233–264). New York: Springer.
Kochanska, G. (1993). Toward a synthesis of parental socialization and child temperament in early development of conscience. *Child Development, 64,* 325–347.
Koestner, R., Weinberger, J., & McClelland, D. C. (1991). Task-intrinsic and social-extrinsic sources of arousal for motives assessed in fantasy and self-report. *Journal of Personality, 59,* 57–82.
Koppitz, E. (1968). *Psychological evaluation of children's human figure drawings.* New York: Grune & Stratton.
Koppitz, E. M. (1984). *Psychological evaluation of human figure drawings by middle school pupils.* New York: Grune & Stratton.
Kuhl, J., & Beckmann, J. (Eds.) (1994). *Volition and personality.* Seattle, WA: Hogrefe & Huber.
Lerner, P. (1992). Toward an experiential psychoanalytic approach to the Rorschach. *Bulletin of the Meninger Clinic, 56,* 451–464.
Leve, R. M. (1995). *Child and adolescent psychotherapy: Process and integration.* Boston: Allyn & Bacon.
Lewin, K. (1935). *A dynamic theory of personality.* New York: McGraw-Hill.
Locraft, C., & Teglasi, H. (1997). Teacher rated empathic behavior and children's T.A.T. stories. *Journal of School Psychology, 35,* 217–237.
Luborsky, L., & Crits-Christoph, P. (1990). *Understanding transference: The CCRT method.* New York: Basic Books.
Lundy, A. (1985). The reliability of the Thematic Apperception Test. *Journal of Personality Assessment, 49,* 141–145.
Machover, K. (1949). *Personality projection in the drawing of the human figure.* Springfield, IL: Charles C. Thomas.
MacLeod, C., & Cohen, I. L. (1993). Anxiety and the interpretation of ambiguity: A text comprehension study. *Journal of Abnormal Psychology, 102,* 238–247.
Maloney, M. P., & Ward, M. P. (1976). *Psychological assessment: A conceptual approach.* New York: Oxford University Press.
McAdams, D. P. (1990). Unity and purpose in human lives: The emergence of identity as a life story. In A. I. Rabin, R. Zucker, R. Emmons, & S. Frank (Eds.), *Studying persons and lives* (pp. 148–200). New York: Springer.
McAdams, D. P. (1995). What do we know when we know a person? *Journal of Personality, 63,* 365–396.
McClelland, D. C., & Koestner, R. (1992). The achievement motive. In C. P. Smith (Ed.), *Motivation and personality: Handbook of thematic content analysis* (pp. 143–152). New York: Cambridge University Press.
McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review, 96,* 690–702.
McGreevy, J. C. (1962). Interlevel disparity and predictive

efficiency. *Journal of Projective Techniques, 26,* 80–87.
Meissner, W. W. (1974). Differentiation and integration of learning and identification in the developmental process. *Annual of Psychoanalysis, 2,* 181–196.
Meissner, W. W. (1981). Internalization in psychoanalysis. *Psychological Issues Monograph, 50.* New York: International Universities Press.
Mendez, M. F., Ala, T., & Underwood, K. L. (1992). Development of scoring criteria for the clock drawing task in Alzheimer's disease. *Journal of the American Geriatric Society, 40,* 1095–1099.
Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.
Moffit, K. H., & Singer, J. A. (1994). Continuity in the life story: Self-defining memories, affect, and approach/avoidance personal strivings. *Journal of Personality, 62,* 21–43.
Morgan, C. O., & Murray, H. A. (1935). A method for investigating fantasies: The Thematic Apperception Test. *Archives of Neurology and Psychiatry, 34,* 289–306.
Murray, H. A. (1938). *Explorations in personality.* New York: Oxford University Press.
Murray, H. A. (1943). *Thematic Apperception Test manual.* Cambridge, MA: Harvard University Press.
Murstein, B. I. (1965). The stimulus. In B. Murstein (Ed.), *Handbook of projective techniques.* New York: Basic Books.
Murstein, B. I. (1968). Efforts of stimulus, background, personality, and scoring system on the manifestation of hostility on the TAT. *Journal of Consulting and Clinical Psychology, 32,* 355–365.
Norcross, J. C. (Ed.) (1986). *Handbook of eclectic psychotherapy.* New York: Brunner/Mazel.
Piaget, J. (1954). *The construction of reality in the child.* New York: Basic Books.
Piotrowski, C. (1984). The status of projective techniques: Or, "wishing won't make it go away." *Journal of Clinical Psychology, 40*(6), 1495–1502.
Piotrowski, C., & Keller, J. W. (1989). Psychological testing in outpatient mental health facilities: A national study. *Professional Psychology: Research and Practice, 20,* 423–425.
Piotrowski, C., Sherry, D., & Keller, J. W. (1985). Psychodiagnostic test usage: A survey of the Society for Personality Assessment. *Journal of Personality Assessment, 49,* 115–119.
Piotrowski, C., & Zalewski, C. (1993). Training in psychodiagnostic testing in APA-approved PsyD and PhD clinical psychology programs. *Journal of Personality Assessment, 61,* 394–405.
Plomin, R. (1986). *Development, genetics, and psychology.* New York: Erlbaum.
Quinn, N., & Holland, D. (1987). *Cultural models in language and thought.* New York: Cambridge University Press.
Rabin, A. I. (1981). Projective methods: A historical introduction. In A. I. Rabin (Ed.), *Assessment with projective techniques* (pp. 1–22). New York: Springer.
Rappaport, D., Gill, M., & Schafer, R. (1968). *Diagnostic psychological testing* (Rev. ed.). New York: International University Press.
Raynor, J. D., & McFarlin, D. B. (1986). Motivation and the self system. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 315–349). New York: Guilford Press.
Ritzler, B. A., Sharkey, K. J., & Chudy, J. F. (1980). A comprehensive projective alternative to the TAT. *Journal of Personality Assessment, 44,* 358–362.
Roback, H. B. (1968). Human figure drawings: Their utility in the clinical psychologist's armamentarium for personality assessment. *Psychological Bulletin, 70,* 1–19.
Rogers, R. (Ed.) (1997). *Clinical assessment of malingering and deception.* New York: Guilford Press.
Rubin, D. C. (Ed.) (1986). *Autobiographical memory.* New York: Cambridge University Press.
Rummelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schematic and sequential thought processes in PDP models. In J. L. McClelland & D. E. Rummelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2). Cambridge, MA: MIT Press.
Sackett, P. R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology, 73,* 482–486.
Saltz, G., & Epstein, S. (1963). Thematic hostility and guilt responses as related to self-reported hostility, guilt, and conflict. *Journal of Abnormal and Social Psychology, 67,* 469–479.
Sandler, J, & Rosenblatt, B. (1962). The concept of the representational world. *The Psychoanalytic Study of the Child, 17,* 128–145.
Schank, R. C. (1990). *Tell me a story: A new look at real and artificial memory.* New York: Charles Scribner's Sons.
Shapiro, D. (1965). *Neurotic styles.* New York: Basic Books.
Shneidman, E. S. (1951). *Thematic test analysis.* New York: Grune and Stratton.
Singer, J. L. (1981). Research applications of projective methods. In A. I. Rabin (Ed.), *Assessment with projective techniques* (pp. 297–331). New York: Springer.
Singer, J. L., & Salovey, P. (1991). Organized knowledge structures and personality. In M. J. Horowitz (Ed.), *Personal schemas and maladaptive interpersonal patterns* (pp. 33–80). Chicago: University of Chicago Press.
Singer, J. A., & Singer, J. L. (1994). Social-cognitive and narrative perspectives on transference. In J. M. Masling & R. F. Bornstein (Eds.), *Empirical perspectives on object relations theory.* Washington, DC: American Psychological Association.
Skinner, B. F. (1953). *Science and human behavior.* New York: Macmillan.
Smith, B. L. (1994). Object relations theory and the integration of empirical and psychoanalytic approaches to Rorschach interpretation. In I. B. Weiner (Ed), *Rorschachiana XIX: Yearbook of the International Rorschach Society* (pp. 61–77). Seattle, WA: Hogrefe & Huber.
Smith, C. P. (Ed.) (1992a). *Motivation and personality: Handbook of thematic content analysis.* New York: Cambridge University Press.
Smith, C. P. (1992b). Reliability issues. In C. P. Smith (Ed.), *Motivation and personality: Handbook of thematic content analysis* (pp. 126–139). New York: Cambridge University Press.
Snow, R. E. (1974). Representative and quasi-representative designs for research on teaching. *Review of Educational Research, 44,* 265–291.
Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement. *Psychological Bulletin, 112,* 140–154.
Stein, D. J., & Young, J. E. (Eds.) (1992). *Cognitive science and clinical disorders.* San Diego, CA: Academic Press.
Stein, M. J. (1955). *The Thematic Apperception Test* (Rev. ed.). Cambridge, MA: Addison-Wesley.
Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.
Stinson, C. H., & Palmer, S. E. (1991). Parallel and distributed processing models of person schemas and psychopathologies. In M. J. Horowitz (Ed.), *Personal schemas and maladaptive interpersonal patterns* (pp. 339–377). Chicago: University of Chicago Press.
Stricker, G., & Healey, B. J. (1990). Projective assessment of object relations: A review of the empirical literature. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2,* 219–230.
Swensen, C. H. (1968). Empirical evaluations of human figure drawings: 1957–1966. *Psychological Bulletin, 70,* 20–44.
Swindell, C. S., Holland, A. L., Fromm, D., & Greenhouse, J. B. (1988). Characteristics of recovery of drawing ability in left and right brain-damaged patients. *Brain and Cognition, 7,* 16–30.
Taylor, S. E., & Crocker, J. (1981). Schematic bases of social information processing. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social Cognition. The Ontario symposium on personality and social psychology.* Hillsdale, NJ: Erlbaum.
Teglasi, H. (1993). *Clinical use of story telling: Emphasizing the TAT with children and adolescents.* Needham Heights, MA: Allyn & Bacon.
Thomas, A. D., & Dudek, S. Z. (1985). Interpersonal affect in Thematic Apperception Test responses: A scoring system. *Journal of Personality Assessment, 49,* 30–36.
Tomkins, S. S. (1947). *Thematic Apperception Test.* New York: Grune & Stratton.
Tomkins, S. S. (1979). Script theory: Differentiated magnification of affects. In H. E. Howe, Jr. & R. A. Dienstbier (Eds.), *Nebraska Symposium on Motivation* (Vol. 26). Lincoln, NE: University of Nebraska Press.
Tomkins, S. S. (1987). Script theory. In J. Aronoff, A. J. Rabin, & R. A. Zucker (Eds.), *The emergence of personality* (pp. 147–216). New York: Springer.
Tomkins, S. S. (1991). *Imagery, affect, consciousness* (Vol. 3). New York: Springer.
Torgesen, J. K. (1977). The role of non-specific factors in task performance of learning disabled children: A theoretical assessment. *Journal of Learning Disabilities, 10,* 27–35.
Tunis, S. L. (1991). Causal explanations in psychotherapy: Evidence for Target-and-Domain-Specific schematic patterns. In M. J. Horowitz (Ed.), *Personal schemas and maladaptive interpersonal patterns* (pp. 261–276). Chicago: University of Chicago Press.
Uleman, J. S., & Bargh, J. A. (Eds.) (1989). *Unintended thought.* New York: Guildford Press.
Urban, H. M. (1963). *The Draw-A-Person.* Los Angeles: Western Psychological Services.
Vane, J. R., & Guarnaccia, V. J. (1989). Personality theory and personality assessment measures: How helpful to the clinician? *Journal of Clinical Psychology, 45,* 5–19.
Veroff, J. (1992). In C. P. Smith (Ed.), *Motivation and personality: Handbook of thematic content analysis* (pp. 100–109). New York: Cambridge University Press.
Watkins, C. E., Campbell, V. L., & McGregor, P. (1988). Counseling psychologists' uses of and opinions about psychological tests: A contemporary perspective. *The Counseling Psychologist, 16,* 476–486.
Watkins, C. E., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology Research and Practice, 26,* 54–60.
Weiner, I. B. (1977). Approaches for Rorschach validation. In M. Rickers-Ovsiankina (Ed.), *Rorschach psychology.* Huntington, NY: Krieger.
Weiner, I. B. (Ed.) (1994). *Rorschchianna XIX: Yearbook of the International Rorschach Society.* Seattle, WA: Hografe & Huber.
Wertheim, E. H., & Schwartz, J. C. (1983). Depression, guilt, and self-management of pleasant and unpleasant events. *Journal of Personality and Social Psychology, 45,* 884–889.
Westen, D. (1991). The clinical assessment of object relations using the TAT. *Journal of Personality Assessment, 56,* 56–74.
Westen, D. (1993). Social cognition and social affect in psychoanalysis and cognitive psychology: From regression analysis to analysis of regression. In J. W. Barron, M. N. Eagle, & D. L. Wolitzky (Eds.), *Interface of psychoanalysis and psychology* (pp. 375–388). Washington, DC: American Psychological Association.
Willock, B. (1992). Projection, transitional phenomena, and the Rorschach. *Journal of Personality Assessment, 59,* 99–116.
Wilson, A. (1988). Levels of depression and clinical assessment. In H. D. Lerner & P. M. Lerner (Eds.), *Primitive mental states and the Rorschach* (pp. 441–462). Madison, CT: International Universities Press.
Wyatt, F. (1947). The scoring and analysis of the Thematic Apperception Test. *Journal of Psychology, 24,* 319–330.
Wyer, R. S., Jr., & Srull, T. K. (Eds.) (1994). *Handbook of social cognition* (2nd ed., Vols. 1 & 2). Hillsdale, NJ: Erlbaum.

4.15
Projective Assessment of Children and Adolescents

IRVING B. WEINER and KATHRYN KUEHNLE
University of South Florida, Tampa, FL, USA

4.15.1 INTRODUCTION 432
4.15.2 OBJECTIVITY AND SUBJECTIVITY IN PERSONALITY ASSESSMENT METHODS 432
4.15.3 STRUCTURE AND AMBIGUITY IN PROJECTIVE TECHNIQUES 433
4.15.4 VALUE OF PROJECTIVE ASSESSMENT 435
4.15.4.1 Conceptual Basis 435
4.15.4.2 Empirical Basis 436
4.15.5 UTILITY OF PROJECTIVE ASSESSMENT 436
4.15.6 APPLICABILITY OF PROJECTIVE METHODS TO CHILDREN AND ADOLESCENTS 437
4.15.7 REVIEW OF PROJECTIVE ASSESSMENT METHODS 439
4.15.7.1 Rorschach Inkblot Method 439
4.15.7.1.1 Administration and scoring 439
4.15.7.1.2 Psychometric foundations 439
4.15.7.1.3 Clinical utility 440
4.15.7.2 Thematic Apperception Test 440
4.15.7.3 Administration and scoring 441
4.15.7.3.1 Psychometric foundations 441
4.15.7.3.2 Clinical utility 442
4.15.7.4 Children's Apperception Test 442
4.15.7.4.1 Administration and scoring 443
4.15.7.4.2 Psychometric foundations 443
4.15.7.4.3 Clinical utility 443
4.15.7.5 Roberts Apperception Test for Children 444
4.15.7.5.1 Administration and scoring 444
4.15.7.5.2 Psychometric foundations 444
4.15.7.5.3 Clinical Utility 445
4.15.7.6 Tell-me-a-story 445
4.15.7.6.1 Administration and scoring 446
4.15.7.6.2 Psychometric foundations 446
4.15.7.6.3 Clinical utility 447
4.15.7.7 Draw-a-person 447
4.15.7.7.1 Administration and scoring 448
4.15.7.7.2 Psychometric foundations 448
4.15.7.7.3 Clinical utility 449
4.15.7.8 House-tree-person 449
4.15.7.8.1 Administration and scoring 450
4.15.7.8.2 Psychometric foundations 450
4.15.7.8.3 Clinical utility 450
4.15.7.9 Kinetic Family Drawing 450
4.15.7.9.1 Administration and scoring 451

4.15.7.9.2 Psychometric foundations 451
4.15.7.9.3 Clinical utility 452
4.15.7.10 Sentence Completion Methods 452
4.15.7.10.1 Administration and scoring 452
4.15.7.10.2 Psychometric foundations 453
4.15.7.10.3 Clinical utility 453
4.15.8 FUTURE DIRECTIONS 454
4.15.9 SUMMARY 454
4.15.10 REFERENCES 455

4.15.1 INTRODUCTION

Projection, as first formulated by Freud (1962) and later elaborated in his presentation of the Schreber case (Freud, 1958), consists of attributing one's own characteristics to external objects or events without adequate justification or conscious awareness of doing so. Frank (1939) suggested that personality tests in which there is relatively little structure induce a subject to "project upon that plastic field ... his private world of personal meanings and feelings" (pp. 395, 402). By linking the concept of projection to the response process in such measures as the Rorschach Inkblot Method and the Thematic Apperception Test (TAT), Frank gave birth to the so-called projective hypothesis in personality assessment. His observations about what he called "projection measures" led to the Rorschach, the TAT, and other assessment methods involving some ambiguity being routinely designated as projective tests.

This chapter begins with some observations concerning the nature of objectivity and subjectivity in personality assessment methods and the role of structure and ambiguity in different types of projective techniques. It then turns to the value and utility of projective assessment and the applicability of projective methods in clinical work with young people. Information is given on the composition, administration, scoring, psychometric foundations, and clinical utility of nine projective techniques widely used in assessing children and adolescents.

4.15.2 OBJECTIVITY AND SUBJECTIVITY IN PERSONALITY ASSESSMENT METHODS

As a legacy of Frank's projective hypothesis, tests designated as projective methods came to be regarded as subjective in nature and hence quite different from objective methods, such as self-report inventories. This presumed subjectivity of projective methods fostered a commonplace conviction that these methods are inherently less scientific and less valid than objective methods. In actuality, however, ambiguity is a dimensional rather than a categorical characteristic of tests, and there is little basis for regarding projective methods as inherently unscientific and invalid or as sharply distinct from objective measures.

Being scientific does not inhere in the nature of a method or instrument, whether subjective or not, but only in whether it can be studied scientifically. When projective tests are used to generate personality descriptions that can be independently and reliably assessed for their accuracy, they function as a scientific procedure.

Likewise, the validity of a test inheres not in its nature, but rather in the extent to which it generates significant correlations with personality characteristics or behaviors it can identify or predict. Abundant research attests that projective methods, when properly used, can yield valid inferences (Hibbard et al., 1994; Hibbard, Hilsenroth, Hibbard, & Nash, 1995; Parker, Hanson, & Hunsley, 1988; Weiner, 1996).

Regarding sharp distinctions between projective and objective measures, the subject's task on commonly used personality tests injects considerable objectivity into many projective methods and substantial subjectivity into most objective methods. For example, responses on the Rorschach, which is the most widely used projective measure, are routinely inquired by asking subjects "Where did you see it?," which is a concrete, unambiguous request for a specific item of objective information. When subjects reply that they used the whole blot for a percept, their response is given a location choice code of W, which is an objective and unambiguous procedure on which coders achieve virtually 100% agreement. Numerous other features of how subjects choose to look at the inkblots can also be objectively coded with good inter-rater reliability, such as the percentage of responses given to the multicolored cards (affective ratio), the number of commonly given percepts reported (populars), and the kinds of objects the blots are said to resemble (people, animals, etc.) (Exner, 1991, pp. 459–460; McDowell & Acklin, 1996; Seaton & Allen, 1996).

Elements of objectivity mark the interpretation as well as the coding of many Rorschach variables. In the case of whole responses, for example, an unusual preponderance of W in a record correlates with objectively observable tendencies to attend to experience in a global fashion; a low affective radio identifies an inclination to withdraw from emotionally charged situations; a small number of popular responses correlates with behavioral manifestations of unconventionality; and numerous human percepts is associated with an active interest in people. In these and many other ways, Rorschach responses identify personality characteristics through an objective process of coding response features and relating these coded features to their known corollaries in observable behavior.

There are aspects of Rorschach interpretation that may be highly subjective, especially when inferences are drawn from the thematic imagery subjects produce when they associate to the inkblots along with describing them. Moreover, most other projective methods have not been as extensively codified as the Rorschach and depend more on qualitative than quantitative analysis. The present point is merely that the basic nature of projective methods does not preclude their being codified and interpreted to some extent along objective lines.

Turning now to aspects of subjectivity in objective methods, consider the uncertainty that characterizes many items in the most widely used objective measure of personality, the adult and adolescent forms of the Minnesota multiphasic personality inventory (MMPI-2/MMPI-A). Although MMPI-2/MMPI-A instructions to respond true or false are unambiguous and the coding of these responses is completely objective, Weiner (1993) has previously called attention to the idiography that is embedded in asking subjects to interpret such items as "I often lose my temper." Items of this type provide no benchmarks for the frequency of "often" or for what constitutes loss of temper. In many instances, consequently, responses to self-report items involve subjectivity on the part of respondents, who must define for themselves what certain terms mean before they can decide how to answer.

Subjectivity influences the interpretation as well as the response process in objective assessment. Granted, the hallmark of objective tests is an extensive array of quantified scale scores having empirically demonstrated behavioral correlates. Nevertheless, the interpretation of self-report measures in clinical practice typically goes beyond identifying known corollaries of scale scores to include consideration of complex patterns of interaction among these scores and between the test profile and aspects of a subject's clinical history, interview behavior, and performance on other tests. Some of these complex interactions have been examined empirically, such as various two- and three-point codes on the MMPI-2/MMPI-A, but many have not.

This is not to say that the MMPI-2/MMPI-A and other self-report instruments are basically subjective in nature or that they derive their utility primarily from clinical judgment. The point is merely that, just as projective instruments are not entirely subjective, self-report methods are not completely objective, but instead involve some aspects of ambiguity in how subjects respond to them and how examiners interpret them.

It is for this reason that ambiguity is not a categorical function that characterizes some tests called projective measures, but not others called objective measures. Instead, ambiguity is a dimensional function that characterizes most tests to some degree, in relation to how structured they are. Generally speaking, objective tests are more structured than projective tests and therefore less ambiguous; projective tests are generally less structured than objective tests and hence more ambiguous; and there is no sharp objective/subjective dichotomy between relatively structured and relatively unstructured instruments.

4.15.3 STRUCTURE AND AMBIGUITY IN PROJECTIVE TECHNIQUES

Projective techniques comprise inkblot methods, story-telling methods, figure drawing methods, and sentence completion methods. In addition to being less structured than objective measures, these four types of projective methods differ from each other in their degree of ambiguity and in whether their ambiguity resides mainly in their stimuli or in their instructions.

Thus in the case of the Rorschach inkblot method, subjects are asked to look at relatively ambiguous stimuli but are given fairly specific instructions to indicate what they see, where they see it, and what makes it look as it does. Story-telling methods such as the TAT involve showing subjects real pictures that are much less ambiguous than inkblots; however, by using general instructions ("Tell me a story") and open-ended questions ("What will happen next?"), examiners provide only minimal guidance in how subjects should respond. If the Rorschach instructions were "Tell me a story about this inkblot," the Rorschach would be more ambiguous and more of a projective test

than it is. If the TAT instructions were "Tell me what you see here," the TAT would lose most of its ambiguity and function only barely as a projective test.

Figure drawing techniques use no stimuli at all, save a blank piece of paper, and provide little guidance to subjects, other than some instructions concerning the figures to be drawn (e.g., yourself, a family). Sentence completion methods, in common with story-telling techniques, call for subjects to provide thematic content in response to real and relatively unambiguous test stimuli. Unlike story-telling techniques, however, sentence completion methods do not ordinarily involve querying subjects about their responses or encouraging them to elaborate those that are brief or unrevealing. Thus a stem of "I AM" may be completed with "a happy person," in which case some subjectivity has been allowed to enter the response, or simply with "here," in which case only a completely objective response has been given. On balance, figure drawing techniques are the most ambiguous of projective tests and sentence completion methods the least, with inkblot and story-telling techniques in between.

These differences in ambiguity among projective methods were originally noted by Stone and Dellis (1960), who proposed "a levels hypothesis" to take practical account of this variability. According to the levels hypothesis, the degree to which a test is structured is directly related to the level of conscious awareness at which it taps personality processes. The more structured and less ambiguous a test is, the more likely it is to yield information about relatively conscious and superficial levels of personality; conversely, the less structured and more ambiguous a test is, the more likely it is to provide information about deeper levels of personality and characteristics of which subjects themselves may not be consciously aware.

Research reported by Stone and Dellis (1960) and subsequently replicated by Murstein and Wolf (1970) provided empirical support for a relationship between the ambiguity of a test and its likelihood of measuring deeper levels of personality, especially in normally functioning persons. These findings mirrored the basic conception of TAT assessment articulated by Murray (1951), who regarded the virtue of the instrument as residing not in its revelations about what subjects are able and willing to say about themselves, but in what it conveys about personality characteristics: "the patient is unwilling to tell or is unable to tell because he is unconscious of them" (p. 577).

In addition to differing from objective tests and from each other in their degree of structure and ambiguity, individual projective measures typically include both relatively objective and relatively subjective elements. As elaborated by Weiner (1977), the objective elements of projective test data involve structural features of the manner in which responses are formulated, whereas the subjective elements consist of thematic features of the imagery with which responses are embellished.

When projective test data are being interpreted objectively, structural aspects of the subject's responses, such as focusing on wholes and seeing numerous human figures on the Rorschach, are taken as being directly representative of similar behavioral tendencies in the person's life, that is, attending to experience globally and paying close attention to people. When projective test data are being interpreted subjectively, thematic imagery is taken as being indirectly symbolic of a subject's underlying needs, attitudes, conflicts, and concerns. Thus the Rorschach response of "Two girls who are really mad at each other fighting over something they both want" may identify a subject's experiencing peer or sibling rivalry, viewing social interactions as aggressive confrontations in which people are only concerned with what they can get for themselves, or feeling angry or resentful about being in such situations.

On story-telling measures, an example of a structural response feature is giving long stories, which can be objectively scored (by counting the number of words) and which provides a representative indication of inclinations to be verbose. As for subjectively interpreted features, a TAT story in which two people are described as about to separate, leaving one of them sad and lonely for the rest of his or her life, exemplifies thematic imagery that appears to symbolize concerns about suffering the loss of love objects and facing an unhappy future.

On figure drawing measures, which as previously noted are the most ambiguous of projective tests, structural features of the data are limited. Some variables, such as the size of figures drawn, how complete they are, and whether they are clothed, are objective facts that can usually be coded with good agreement. However, interpretation of such objective characteristics of figure drawings, as well as of subjective impressions of drawing qualities, is based mostly on their being symbolic rather than representative of behavior. Interpreting the way figures are drawn or placed is thus primarily thematic. For example, unusual emphasis on a particular part of the body may be interpreted as suggesting concern about functions associated with that part of the body, and a family drawing in which the self is located on one side of the page and the other family members are closely grouped on the other side of the page may be interpreted as symbolizing feelings of isolation or rejection in the family setting.

In sentence completion responses, frequent self-referencing is an example of an objectively scorable, behaviorally representative structural index of tendencies to focus attention on oneself rather than others. Consider the difference between the completions "WHAT PAINS ME is seeing how many unfortunate people there are in the world" and "WHAT PAINS ME is not being able to get the things that I want." An accumulation of the latter as opposed to the former type of response is objectively representative of self-centeredness. At the same time, the thematic content of both completions suggests in a more subjective way certain underlying concerns, such as worries about the welfare of the human race in the first instances and feelings of being personally deprived in the second.

To bring these introductory observations full circle, the opportunities that projective methods create for subjects to project aspects of themselves into their responses has frequently led to their being associated with psychoanalytic theories of personality, in the context of which the notion of projection was first elaborated. However, there is no necessary relationship between psychoanalytic theory and projective testing, nor is there any reason for clinicians who conceptualize behavior in other ways to view projective methods as incompatible with their frame of reference. The basic principle underlying projective techniques is that something can be learned about people from sampling how they respond in ambiguous situations. This principle is not prisoner to any personality theory, and its utility transcends the theoretical persuasions of individual examiners. Inferences from projective data can be couched equally well in psychodynamic, behavioral, cognitive, and humanistic terms, and the use to which these inferences can be put depends less on theoretical differences in terminology than on the nature of the assessment issues being addressed.

4.15.4 VALUE OF PROJECTIVE ASSESSMENT

Projective test data provide valuable information about how people are likely to think, feel, and act that is difficult to obtain from objective assessment procedures. This contribution of projective methods to the personality assessment process has both a conceptual and an empirical basis.

4.15.4.1 Conceptual Basis

Because of their relatively unstructured nature, projective tests measure personality characteristics in subtle and indirect ways. Even those features of projective test data that can be objectively scored and interpreted involve responses that seldom have obvious meaning. Subjects in the process of responding usually have little awareness of the interpretive significance that attaches to their seeing numerous human figures on the Rorschach, giving long stories on the TAT, drawing themselves on the far side of the page from the rest of their family, or repetitively referring to "I" in their sentence completions; indeed, they may not even be aware of having responded in these ways.

By contrast, relatively structured objective tests measure personality characteristics in direct ways that often have obvious interpretive significance. Adolescents who answer "true" to such MMPI-A statements as "At times I feel like smashing things" and "I am easily downed in an argument" will usually have a good idea of what they are indicating about themselves.

The distinction between subtle, indirect measurement of personality characteristics with projective techniques and relatively direct assessment through questionnaire methods has been formulated by McClelland, Koestner, and Weinberger (1989) in terms of differences between self-attributed and implicit motives. According to McClelland et al., self-attributed motives are measured by self-report instruments and are influenced by social incentives in a person's external environment. Implicit motives, however, are measured by such indirect techniques as story-telling procedures and are influenced by the internal pleasure derived from various activities in which a person engages.

Whereas self-attributed motives are comparatively good predictors of immediate specific responses to structured situations, McClelland et al. continue, implicit motives are comparatively good predictors of long-term trends in behavior across various types of situations. Research findings described by McClelland et al. confirmed that indirect assessments of underlying motives have greater validity for predicting long-term trends in behavior than self-report assessments of motives that people directly attribute to themselves.

More recently, Bornstein (1995) has used a metaanalysis of 97 studies of measures of dependency to demonstrate further this difference between objective and projective assessment. With respect to differential prediction, according to Bornstein, available research indicates that objectively measured dependency correlates better with symptoms and the

diagnosis of dependent personality disorder than does projectively measured dependency, whereas projectively measured dependency correlates better with dependency-related behaviors.

The conceptual analysis formulated by McClelland et al. and elaborated by Bornstein has direct bearing on the contribution of projective methods to personality assessment. As previously described, projective assessment taps implicit motives and underlying personality characteristics that may not be readily apparent and may not be within a subject's conscious awareness. Because these covert motives and characteristics exert a powerful influence on long-term behavioral trends, this type of indirect measurement adds an important dimension to personality evaluations that would not be tapped in its absence.

Finally, with respect to what projective methods contribute to assessment batteries, the relative ambiguity of these methods makes them less subjective than structured instruments to influence by test-taking attitudes. This is not to say that projective methods are immune to subjects' efforts to present themselves in a positive or negative light. The relatively open-ended nature of projective testing situations and the dialogue they frequently elicit give subjects abundant opportunity to voice attitudes toward the tests, the examiner, and being examined. However, as long as subjects continue to give responses, neither their attitudes nor their expression of them is likely to prevent their projective test responses from revealing their personality characteristics. Simply put, the limited face validity of projective measures makes them more difficult to fake than objective measures, which means that they can balance a test battery to particularly good effect when self-presentation effects are of concern.

4.15.4.2 Empirical Basis

Projective measures vary in the extent to which they have been examined in well-designed research studies, and in many instances adequate empirical support for these instruments has lagged behind the uses to which clinicians sometimes put them. Nevertheless, the two most frequently used projective methods, the Rorschach and the TAT, have for the past generation been among the three most frequently studied personality assessment instruments, exceeded in this respect only by the MMPI/MMPI-2 (Butcher & Rouse, 1996). For both the Rorschach and the TAT, substantial evidence has accumulated to attest their validity for describing aspects of personality structure and dynamics and applying these descriptions in differential diagnosis and treatment planning (Abraham, Lepisto, Lewis, & Schultz, 1994; Alvarado, 1994; Bornstein, 1995; Cramer & Blatt, 1990; Exner & Andronikoff-Sanglade, 1992; Ornduff & Kelsey, 1996; Ronan, Colavito, & Hammontree, 1993; Weiner, 1996; Weiner & Exner, 1991).

Later in this chapter, specific information is presented concerning the psychometric foundations and demonstrated corollaries of the projective measures most frequently used with young people. Suffice it to say in summary at this point that these measures prove valuable in personality assessment because they add information that would otherwise be unavailable and because they withstand relatively well efforts to exaggerate or conceal.

4.15.5 UTILITY OF PROJECTIVE ASSESSMENT

Whereas the value of projective techniques lies in the previously elaborated reasons *why* they should be used in personality assessment, their utility relates to decisions concerning *when* these methods should be included in a test battery. The kinds of information provided by projective test data indicate that projective measures should be used whenever a thorough personality assessment is considered relevant to formulating a differential psychodiagnosis or recommending alternative intervention strategies. Because of their relatively unstructured nature and indirect format, projective measures balance a test battery by tapping personality characteristics at a less conscious level than relatively structured measures. Assessments lacking such balance sample personality functioning from a limited perspective that will usually fail to paint a complete picture of the individual being examined. Batteries limited solely to projective techniques are similarly imbalanced and ill-advised in comprehensive personality assessments.

The previously mentioned conceptualization of McClelland et al. (1989) bears closely on the importance of a balanced test battery in clinical assessment. McClelland and his colleagues noted that measures of self-attributed and implicit motives seldom correlate with each other and should not be expected to do so, because they are measuring different aspects of personality. Moreover, given that directly and indirectly measured motives each predict certain kinds of behavior better than the other, they concluded that "Separate measures of self-attributed and implicit motives may be combined to yield a better understanding and prediction of certain types of behavior" (p. 692).

These formulations concerning different types of measures have subsequently been elaborated and confirmed for clinical purposes with respect to relationships between the Rorschach and the MMPI. Rorschach structural variables and MMPI scales have been found to show only a few modest correlations in both adult and adolescent samples (Archer & Krishnamurthy, 1993a, 1993b). At the same time, however, apparent contradictions between Rorschach and MMPI findings have been conceptualized by Weiner (1993, 1995b) not as invalidating either instrument or challenging the incremental utility of administering them in tandem, but rather as generative data.

Specifically, Weiner argues, apparently discrepant findings between personality assessment instruments of different kinds can be generative by virtue of complementing each other. Whereas findings on two tests that concur in suggesting the same personality characteristic are confirmatory and support definite conclusions, he continues, findings that diverge raise important questions to which they may also suggest helpful answers, especially if one of the tests is a relatively structured and the other a relatively unstructured instrument.

Consider, for example, a youngster who appears depressed on the Rorschach, with a high depression index (DEPI), but does not elevate on Scale 2 or the depression content scale of the MMPI-A. This divergent finding could well provide a useful clue to the adolescent's having an underlying or emerging depression that is not yet being keenly felt or manifest in well-structured situations. Alternatively, it could be that the subject is trying to deny or repress depressive affects and cognitions, or is making a conscious decision not to report manifestations of depression that nevertheless emerge in the absence of supportive structure or are revealed when the subject is uncertain how to conceal them.

The situation described in this example is familiar to assessment psychologists, who not infrequently work with psychologically troubled adolescents who can remain reasonably comfortable and controlled in relatively structured situations but become upset and disorganized in relatively unstructured situations and who may accordingly produce a benign MMPI-A protocol and a disturbed Rorschach. In such circumstances the objective measure has not erroneously overlooked psychopathology, nor has the projective measure mistakenly exaggerated it. Instead, the two types of test have combined in complementary fashion to provide valid information concerning the subject's likelihood of behaving in a relatively adaptive or maladaptive fashion, depending on the situational context in which the behavior appears.

Similarly, subjects in some circumstances may produce a clinically unremarkable Rorschach while showing numerous elevations on the clinical and content scales of the MMPI-A. Such divergence is best understood not as error variance, but as a possible clue to the psychological stance of subjects whose degree of disturbance is minimal but who, when asked about themselves in language they can understand, want to make sure that others fully appreciate whatever problems and concerns they do have. Further illustrations of the clinical utility of divergence as well as convergence between a projective measure, such as the Rorschach and an objective measure such as the MMPI-2, are provided by Finn (1996) and Ganellen (1996).

4.15.6 APPLICABILITY OF PROJECTIVE METHODS TO CHILDREN AND ADOLESCENTS

Except for an occasional example, this chapter has thus far made no specific reference to young people. This apparent oversight is warranted by the fact that the nature of projective methods, the way in which they function, the kinds of information they provide, and the reasons for using them, are identical for persons of almost all ages. Hence the discussion of projective assessment to this point is as applicable to children and adolescents as to adults, and requires no modification or qualification as our focus now shifts specifically to young people.

Indeed, assessors who have learned to interpret projective test data provided by adults do not need to learn any new ways of working with the data should they begin to examine children and adolescents. By and large, the basic interpretive conclusions and hypotheses that attach to projective test variables apply regardless of the age of the subject. Whether they are age 8, 18, or 80, subjects who see numerous human figures on the Rorschach are likely to be quite interested in people; those who give long TAT stories are likely to be verbose; those who refer frequently to themselves in sentence completions are likely to be self-centered; and those who draw grotesquely distorted human figures probably harbor some disturbing concerns about their own nature or that of other people.

However, in order to determine the implications of these and other personality characteristics suggested by projective test data, examiners assessing young people must take

into account normative developmental expectations. For example, the data of developmental psychology indicate that children are more self-centered than adults, and subsequently become increasingly aware of and concerned about the needs of others as they grow through adolescence and approach maturity. Accordingly, test data that identify a high degree of self-centeredness may imply maladaptive narcissistic personality traits in an adult, but reflect normal development and adaptation in a child; conversely, minimal self-centeredness may indicate altruism and good adjustment in an adult but suggest deviant development and low self-esteem in a child.

Developmental psychology similarly provides some normative expectations for how children are likely to make drawings. Preschool age children commonly draw with what is called "intellectual realism," which means that they draw what they know to be there regardless of whether it would actually be visible. Thus, in x-ray fashion, young children often draw transparencies, such as people who are visible through walls (Di Leo, 1983). At about age 7 or 8, this intellectual realism gradually gives way to "visual realism," in which what is drawn resembles what realistically can be seen. Di Leo (1983, p. 38) observes that this developmental shift mirrors a metamorphosis in thinking from an egocentric to an increasingly objective view of the world. Hence a human figure drawing by a preschool child showing a belly button through clothing is much less likely to imply maladaptive functioning that the same drawing done by an adolescent.

As these examples indicate, familiarity with and adequate attention to normative expectation hold the key to valid and useful applications of projective methods in the assessment of young people. Ideally, projective methods manuals should include normative reference data that delineate quantitative as well as qualitative expectations for such developmental phenomena as maturational changes in self-centeredness. Regrettably, even though numerous projective test variables have been quantified in various ways, little progress has been made in generating age-graded norms for them.

The main exception to this dearth of normative developmental data for projective techniques is the Rorschach. Developmental trends in Rorschach responses from early childhood through adolescence were initially charted many years ago by Ames and her colleagues (Ames, Metraux, Rodell, & Walker, 1974; Ames Metraux, & Walker, 1971). More recently, the Rorschach comprehensive system has provided reference data for each of its codified variables on samples of approximately 100 nonpatient young people at each age from 5 to 16 (Exner & Weiner, 1995, chap. 3).

Aside from identifying needs for further research, an analysis of available data can guide clinicians in choosing which projective methods to include in a battery for assessing a young person's functioning. The more thorough and reliable the normative developmental data available for the instrument, the better the choice it will make. Similarly, the better established an instrument's correlates are in relation to behaviors that are central to the purpose of an assessment, the more reason there is to include it. Thus an instrument that has been demonstrated to be particularly helpful in identifying youthful depression may be a good choice in one case, whereas an instrument known to be especially sensitive in revealing family dynamics may be a good choice in another case.

Similarly, available empirical data and reported clinical experience should be drawn on to determine whether a particular instrument is likely to yield useful information concerning the personality functioning of individuals at certain ages. Thus the children's apperception test (CAT) depicting animal figures may be a more effective story-telling measure for a young child than the TAT, but certainly not for an adolescent (Bellak, 1993, p. 237).

As these observations indicate, projective methods provide sound clinical data only if they are employed in appropriate ways. First, examiners should have recourse to standardized procedures for administering and scoring any test they use. Lack of such standard methodology compromises the value of the data obtained, and inattention to standardized methods by examiners who opt instead for personalized approaches to administration and scoring is clinically disadvantageous and professionally questionable.

Second, clinical interpretations should be derived from test variables with demonstrated reliability and validity. Inadequate psychometric foundations limit the use to which test data can be put, and examiners who draw conclusions in the absence of supporting empirical evidence, without framing such conclusions as speculative hypotheses, are doing their patients and their methods a disservice. Third, the adequacy of projective assessment of young people will be limited in the absence of normative reference data for test responses of both adjusted and maladjusted children and adolescents and for developmental changes in these responses over time.

This chapter continues with reviews of the major inkblot, story telling, figure drawing, and sentence completion methods used in assessing young people. The composition, administration, and scoring of each of these measures are described; what is known about their reliability, validity, and normative database is reported; and the clinical purposes they are likely to serve are discussed. The specific measures reviewed are selected primarily on the basis of their emphasis and frequency of use in clinical and school settings, as reported in surveys by Archer, Imhof, Maruish, and Piotrowski (1991), Elbert and Holden (1987), Hutton, Dubes, and Muir (1992), Kennedy, Faust, Willis, and Piotrowski (1994), Piotrowski and Keller (1989), Stinnett, Havey, and Oehler-Stinnett (1994), and Watkins, Campbell, Nieberding, and Hallmark (1995).

4.15.7 REVIEW OF PROJECTIVE ASSESSMENT METHODS

4.15.7.1 Rorschach Inkblot Method

The Rorschach inkblot method comprises 10 cards that are inked in shades of black and gray (five cards); black, gray, and red (two cards); and various pastel colors (three cards). The cards are reproduced in standard fashion, but the inkblot stimuli were originally designed at random and do not portray any specific objects (Rorschach, 1942). When subjects respond to the Rorschach, they draw on the shape, shading, and color of the blots to form impressions of what they might be, and in so doing they treat the instrument as a cognitive–perceptual task (e.g., "It looks like a bat, because it's got a body here and wings here and it's black"). In addition, subjects frequently elaborate their responses beyond the stimulus properties of the blots, and in so doing they treat the instrument as an associational task (e.g., "This bird is flying around looking for something to eat").

The cognitive–perceptual aspects of responses constitute structural data in Rorschach assessment and provide representative indications of the resources and coping style that a person generally brings to bear in problem-solving situations. The associational aspects of responses constitute thematic data in Rorschach assessment and provide symbolic clues to the underlying needs, attitudes, conflicts, and concerns that are likely to influence a person's actions and state of mind. The basic nature of the Rorschach in these respects is discussed further by Exner and Weiner (1995, chap. 1) and Weiner (1986, 1994).

4.15.7.1.1 Administration and scoring

The Rorschach is introduced by telling subjects that the inkblots they are about to see are not anything in particular, but that people see many different things in them and that their task will be to indicate what the inkblots look like to them. The 10 cards are then given to subjects one at a time with the instruction "What might this be?" Requests for structure (e.g., "Can I turn the card?") are deflected back to the subject (e.g., "It's up to you"; "Any way you like"). The unguided responses to the 10 cards constitute the free association phase of the administration, following which there is an inquiry phase in which the examiner reads back each response and asks subjects where they saw it and what made it look as it did. The purpose of the inquiry is to facilitate coding of the structural features of the responses, and associations during this phase are not requested or encouraged. Responses are recorded verbatim, however, and the content of any spontaneous thematic elaborations is carefully noted.

Numerous approaches to codifying Rorschach responses have emerged during the long history of this instrument. However, for many years the comprehensive system of Exner (1993) has been by far the most widely used and researched (Piotrowski, 1996). Rorschach responses are coded in the comprehensive system for various aspects of where percepts are seen (location), why they look as they do (determinants), what they consist of (content), how commonly they occur (form level and populars), and whether they involve pairs of objects, organization of parts, or special kinds of elaborations, such as cooperative or aggressive interaction. These codes are than tallied and combined in various ways to yield a large number of indices, ratios, and percentages that guide the interpretive process, as elaborated in detail by Exner (1991, chaps 5–10).

4.15.7.1.2 Psychometric foundations

The psychometric foundations of an assessment instrument comprise the extent to which it can demonstrate adequate interscorer agreement, reliable measurement, valid correlates, and a representative normative database. The Rorschach inkblot method, as already indicated in part by examples used earlier in the chapter, rests on a solid psychometric basis. Interscorer agreement for the types of variables coded in the comprehensive system typically ranges from 80% to 100%. The reliability of Rorschach data has been demonstrated in a series of retest studies conducted over intervals ranging from seven days to three years and involving child, adolescent, and adult subjects. Most of the core variables associated with trait dimensions of personality show stability coefficients greater

than 0.80 in these studies, and some, including the affective ratio and the egocentricity index, consistently hover around 0.90 (Exner, 1991, pp. 459–460; Exner & Weiner, 1995, pp. 21–27; McDowell & Acklin, 1996; Weiner, 1997).

The validity of Rorschach assessment was confirmed in a series of metaanalytic studies that led Parker, Hanson, and Hunsley (1988) to conclude that the Rorschach meets usual psychometric standards for validity and is comparable to the MMPI in this respect. Specifically, Parker et al. used the effect sizes reported in 411 studies to derive population estimates of convergent validity of 0.41 for the Rorschach and 0.46 for the MMPI. Subsequent further confirmations of the validity of this instrument are noted by Weiner (1996).

With respect to its normative database, available information for the comprehensive system includes data on 700 nonpatient adults demographically representative of the 1980 US census, 1390 nonpatient children and adolescents age 5 to 16, and large samples of schizophrenic, depressed, and character disordered patients (Exner, 1993, chap. 12). In addition, longitudinal data reported by Exner, Thomas, and Mason (1985) on a group of young people tested every two years from age 8 to 16 provide useful reference information concerning developmental stability and change in Rorschach variables during childhood and adolescence.

As implied by the nature of the normative data, the Rorschach comprehensive system is applicable to young people from age five. Preschool age children have ordinarily not yet matured sufficiently to deal with the cognitive–perceptual aspects of the Rorschach situation in ways that lend themselves to the codification that is central to the comprehensive system interpretive process. An unusually mature four-year-old might on occasion produce a useful record, and immature five- and six-year-olds may produce records that have limited interpretive significance within the framework of the comprehensive system. Working within other frameworks, Ames et al. (1974) discuss and provide some normative findings for Rorschach responses of young children, and Leichtman (1996) has recently presented a developmental rationale for deriving information from the records of preschoolers.

4.15.7.1.3 Clinical utility

In common with projective techniques in general, the Rorschach serves clinical purposes primarily as a result of the information it provides about the structure and dynamics of personality functioning. With respect to personality structure, the Rorschach has proved especially helpful in identifying and quantifying states of subjectively felt distress that combine elements of anxiety and depression and in reflecting trait dimensions of how people typically think, process information, handle emotions, manage stress, feel about themselves, and relate to others. Regarding personality dynamics, the thematic content of Rorschach responses, as previously noted, is often quite revealing of underlying needs, attitudes, conflicts, and concerns that influence how people are likely to think, feel, and act at particular points in time and in particular situations.

In addition, Rorschach data can frequently contribute to differential diagnosis in clinical settings. The comprehensive system provides indices for schizophrenia and depression (DEPI) that can help to identify these conditions in children and adolescents as well as in adults; for basic deficits in coping capacity that point to developmental arrest in young people; and for numerous features of conduct and anxiety and/or withdrawal disorders (Exner & Weiner, 1995, chaps 5–8; Weiner, 1986). Rorschach findings have also demonstrated considerable clinical utility in the treatment process by clarifying treatment targets, identifying potential obstacles to progress in therapy, and providing a basis for evaluating treatment change and outcome (Abraham et al., 1994; Weiner, 1994).

4.15.7.2 Thematic Apperception Test

The most widely known and used story telling technique is the TAT. It was developed by Morgan and Murray (1935) in the belief that the content of imagined stories would provide clues to the underlying dynamics of a subject's interpersonal relationships and self-attitudes. As elaborated by Murray (1943, 1971) and Bellak (1993, chap. 4), TAT data are expected to reveal the hierarchy of a person's needs and the nature of his or her dominant emotions and conflicts.

The TAT stimuli comprise 19 black-and-white illustrations of people or scenes and one blank card. The cards are intended for use with persons age five or older of both genders, and for nine of the cards there are alternate versions for use with adult and child/adolescent males and with adult and child/adolescent females. Because of the time required to administer the full set of TAT cards, examiners typically select a subset of 8–12 cards that they anticipate will elicit themes relevant to the assessment issues in a particular case. The themes usually elicited by the individual cards and the selection of subsets suited for children and adolescents are reviewed by Bellak (1993, chap. 3), Dana (1985), and Obrzut and Boliek (1986). Regrettably with respect to standardization, however, there are no specific short forms of the instrument, and how many and which cards are typically chosen vary from one examiner to another and from one examination to the next.

4.15.7.3 Administration and scoring

The TAT cards are given to subjects one at a time with instructions to make up a story for each picture that includes (i) what is happening at the moment, (ii) what the characters are thinking and feeling, (iii) what led up to the situation, and (iv) what the outcome will be. The narrated stories are recorded verbatim by the examiner.

Murray (1943) originally proposed a scoring scheme in which each TAT story is rated for the presence and strength of a long list of needs that are being experienced by the central figure in the story and presses that are being exerted by the environment. This scoring system proved too elaborate and time consuming for clinical work, and numerous alternative approaches to clinical interpretation were subsequently developed for the instrument. As reviewed by Chandler (1990), Murstein (1963), and Vane (1981), some of these interpretive approaches, like Murray's, have consisted of formal quantitative ratings of story characteristics. However, most interpretive approaches have eschewed quantitative scoring in favor of qualitative analyses of story content, and no scoring system has gained widespread use either clinically or in research studies.

The most commonly employed methods of interpreting the TAT in clinical practice appear to be variations of an "inspection technique" proposed by Bellak (1993, chap. 4). This technique consists simply of reading through subjects' stories to identify repetitive themes and recurring elements that appear to fall together in meaningful ways. Because this approach lacks any quantification and rests on the capacity of individual examiners to relate story themes and elements to aspects of personality functioning, Dana (1985) was moved to observe that "TAT interpretation has become a clinical art form" (p. 90).

Bellak's influential approach stresses 10 aspects of a story, each of which is taken to have implications for how subjects view and are likely to deal with interpersonal events and what they anticipate the future to hold for them. These include the main theme of the story, the identity of the central figure or hero, the main needs of the hero, the way in which the environment is conceived, the identity and intentions of other figures in the story, the nature of any anxiety or other affect that is being experienced, the nature of any conflict that is described or suggested, the ways in which conflicts and fears are defended against, the ways in which misbehavior is punished, and the level of ego integration.

With respect to research studies, the most productive utilization of the TAT has derived from quantitative scoring systems developed by McClelland, Atkinson, and their colleagues to measure needs for achievement, affiliation, and power (Atkinson & Feather, 1966; McClelland, Atkinson, Clark, & Lowell, 1953). Although scoring for achievement, affiliation, and power motivation has had little clinical impact, other schemes for coding specific personality characteristics reflected in TAT thematic content have subsequently emerged. These include scales for level of ego development (Sutton & Swenson, 1983), preferred defense mechanisms (Cramer, 1987), quality of interpersonal affect (Thomas & Dudek, 1985), problem-solving style (Ronan et al., 1993), and object relations capacities (Westen, Lohr, Silk, Kerber, & Goodrich, 1985).

Particularly promising among these is the use of the TAT to assess aspects of object relatedness through the Westen et al. (1985) measure, known as the social cognition and object relations scale (SCORS). The SCORS provides quantitative indices of the affective tone subjects ascribe to relationships, their capacity for emotional investment in relationships and social standards, their understanding of social causality, and the complexity of their representations of people. By including ratings of subjects along dimensions of maturity as well as normality/pathology, the SCORS is proving especially relevant to the assessment of young people (Westen et al., 1991).

4.15.7.3.1 Psychometric foundations

Efforts to demonstrate the reliability and validity of global approaches to interpreting the TAT have been handicapped by the previously noted proliferation of scoring systems, by clinicians' preferences for a strictly qualitative and uncoded approach to the data, and by enormous variation in how the test is administered, including which subset of cards is selected for use. Because of this long-standing lack of standardization, there has been little opportunity for systematic accumulation of data bearing on the reliability and validity of the TAT in general, nor has it been possible to develop any substantial normative database.

Accordingly, for both inspection techniques and overall scoring systems developed in the tradition of Murray, the psychometric literature on the TAT is generally acknowledged to comprise a mix of positive and negative findings that cannot easily be compared with one another. Hence, despite the widespread use of inspection techniques in clinical practice, neither these nor other global approaches have been demonstrated to show adequate psychometric properties.

However, research studies with TAT scales developed to measure specific personality characteristics have demonstrated that the instrument can generate reliable and valid findings when it is used in a standardized manner. The previously mentioned scales of Cramer (1987) and Westen et al. (1985) are cases in point. Cramer's scale reliably identifies three major mechanisms of defense—denial, projection, and identification—and has shown valid corollaries in changes observed in patients undergoing psychotherapy (Cramer & Blatt, 1990). The Westen et al. SCORS has been found to provide reliable identification of developmental variables related to disturbed object relations in children and, as already mentioned, is therefore especially relevant to the assessment of young people. Validation studies with SCORS have involved psychiatrically disturbed, borderline, physically abused, and sexually abused young people (Freedenfeld, Ornduff, & Kelsey, 1995; Ornduff, Freedenfeld, Kelsey, & Critelli, 1994; Westen, Ludolph, Block, Wixom, & Wiss, 1990; Westen, Ludolph, Lerner, Ruffins, & Wiss, 1990).

4.15.7.3.2 Clinical utility

The clinical utility of the TAT lies mainly in its potential for elucidating dynamic aspects of personality functioning, particularly with respect to the feelings and attitudes that subjects hold towards other people, themselves, and possible turns of fortune in their lives for better or worse. Based on the assumption that children and adolescents identify with the central figures in their TAT stories and project fantasies and realities regarding their own lives into the events and circumstances they describe, the obtained data can shed light on a broad range of underlying influences on how young people are likely to think, feel, and act.

As previously noted in commenting on the research of McClelland et al. (1989), the implicit types of motives measured by the TAT are more likely to correlate with persistent dispositions to behave in certain ways rather than with immediate actions or symptom formation. Accordingly, TAT findings will usually not add very much to structural diagnosis of adjustment problems in young people, but they can be extremely helpful in suggesting possible dynamic origins of adjustment problems. In this regard, the psychometrically sound SCORS may be a useful scale to include in forensic assessment batteries when issues of custody or adoption are being addressed. This TAT scale can frequently assist examiners in grasping a young person's representations of people and his or her capacities for emotional investment in relationships.

The thematic content of TAT stories has additional potential to facilitate planning and conducting psychotherapy with young people, particularly with respect to identifying treatment targets and monitoring progress in therapy. The TAT can also be used in treatment as a play therapy tool, as in Gardner's (1971) story telling technique. For example, after a youngster has told TAT stories, the therapist and child can act out the stories in play, or the therapist can create stories to the same picture stimuli for comparisons with the child's stories. Hoffman and Kupperman (1990) describe such an intervention with a 13-year-old boy in which both therapists wrote stories to the same TAT cards to which the patient had responded. As it turned out, Hoffman's stories emphasized the main character's maladaptive coping mechanisms, whereas Kupperman's stories emphasized positive and healthy aspects of the central character's coping capacities. Over a number of sessions, this boy and his therapist engaged in discussions concerning whose version of the story was most accurate.

4.15.7.4 Children's Apperception Test

Consistent with the purpose of the TAT, Bellak (1993, chap. 13) developed the CAT to facilitate understanding of personality processes in children, including their "dynamic way of reacting to and handling the problems of growth" (Bellak & Siegel, 1989, p. 102). The CAT pictures were designed to elicit fantasies about aggression, sibling rivalry, fears of being alone at night, attitudes toward parental figures, and eating problems.

The CAT-Animal (CAT-A) form, originally published in 1949 and designed for children 3–10 years old, consists of 10 pictures depicting animals in human situations. The use of animal figures was based on the assumption that young children identify more readily with animals than with people and will accordingly tell more meaningful stories about animal than human figures. Moreover, according to Bellak (1993, chap. 13), the use of animal figures makes the CAT-A a culture-free test that is equally applicable to Caucasian, African-American, and other minority group youngsters as well as to children from different countries, except where there is little familiarity with some of the inanimate objects depicted, such as bicycles.

There is also a human form of the CAT (the CAT-H) that was developed by Bellak and Hurvich (1966) in response to criticism of the assumption that children identify more easily with animal than with human figures. Studies reviewed by Bellak and Hurvich indicate little difference in stimulus value between the original CAT-A and the CAT-H, in which human figures are substituted for the animals in the CAT-A scenes. However, the CAT-H does not appear ever to have become much used in clinical practice.

4.15.7.4.1 Administration and scoring

Children being administered the CAT are told that they are going to take part in a game in which they will tell stories about pictures. Subjects who appear to regard the CAT as a test are informed that it is not the type of test in which they will be graded for correct or incorrect answers. Standard procedures call for all 10 CAT pictures to be administered in numerical order, from Card 1 to Card 10. Children are told to narrate what the animals are doing in the pictured scenes and are asked at appropriate points to say what went on previously and what will happen next. The examiner encourages and prompts the subject as necessary but avoids being suggestive or asking leading questions. Examiners may also query each story by asking the child to elaborate specific points such as the ages of characters and why they were given particular names.

In clinical work the CAT is typically interpreted along the lines proposed by Bellak for the TAT, that is, with an inspection technique used to form qualitative impressions of various dimensions of the subject's personality functioning (Bellak, 1993, chap. 14). There is an alternative but rarely used quantitative approach developed by Haworth (1965), called the schedule of adaptive mechanisms, in which CAT responses are rated numerically for the degree of adaptability or disturbance they reflect. Haworth also stressed the importance of recognizing that young children are highly reactive to the immediate circumstances in their lives and less likely than adolescents or adults to have formed well-established personality traits. She accordingly emphasized careful interpretation of CAT responses in the context of adequate information concerning subjects' home situation and the nature of any recent or impending crises in their lives.

4.15.7.4.2 Psychometric foundations

There has regrettably been little accumulation of empirical data bearing on the reliability and validity of the CAT. The widespread use of Bellak's qualitative inspection technique in CAT interpretation and a corresponding lack of quantification have precluded examination of the instrument's psychometric foundations. Although it has sometimes been suggested that the idiographic nature of CAT as well as TAT data makes traditional psychometric criteria difficult to apply or even irrelevant, there is nothing in the nature of the data generated by story telling techniques that prevents their being reliably coded for various types of feelings, motives, attitudes, and capacities that can in turn be validated against meaningful correlates. The previously noted development of psychometrically sound TAT scales for such specific aspects of personality as achievement motivation and social cognition proves the point that clinical interpretation of stories can go beyond being an art form and attain respectability as a scientific procedure as well.

Regarding what research is available concerning the CAT, Bellak (1993, chap. 16) provides a review of studies comparing the responses typical of children at different ages and examining special features in the stories of maladjusted, schizophrenic, speech disordered, retarded, brain-damaged, and chronically ill children. Almost all of these studies date from the 1950s, however, and none provides an adequate basis for developing any formal normative standards or diagnostic guidelines for the instrument.

4.15.7.4.3 Clinical utility

As in the case of the TAT, the CAT is useful in clinical assessment primarily as a source of hypotheses concerning subjects' personality dynamics, particularly with respect to how they view themselves and other important people in their lives, the nature of their hopes and fears, and what they expect will happen to them. Although not suitable for adolescents, the CAT is more useful than the TAT in work with younger children, who are likely to relate more easily to the familiar situations and youthful figures depicted in the CAT illustrations than to the primarily adult figures and unpopulated scenes shown in the TAT.

Like the TAT, the CAT may also contribute to treatment planning, by suggesting areas of concern on which to focus in the therapy, and it

may itself serve as a play technique. In diagnostic assessments, however, hypotheses generated by this instrument require support from other data prior to being addressed to questions that necessitate empirical decision making.

4.15.7.5 Roberts Apperception Test for Children

The Roberts apperception test for children (RATC) is intended for use with young people of ages 6–15 and was designed to improve on the TAT and CAT by presenting familiar stimuli and employing a standardized scoring system (McArthur & Roberts, 1990).

Instead of illustrations primarily of adults, such as those used in the TAT, or illustrations of animals, as used in the CAT, the RATC primarily portrays child and adolescent figures engaged in everyday interactions, including scenes of parental affection, disagreement, school and peer relationships, and observation of nudity. There are 27 RATC cards, 11 of which are alternate versions for male or female subjects. There is an addition an alternate set of cards portraying African-American individuals in similar scenes.

4.15.7.5.1 Administration and scoring

The standard 16 cards that compose the RATC are administered individually in numerical order, using male or female versions as appropriate. Subjects are instructed to make up a story about each picture and to tell what is happening in the picture, what led up to the scene, how the story ends, and what the people are talking about and feeling. Responses are recorded verbatim.

If subjects tell an incomplete story or omit certain aspects, such as how the characters are feeling, additional inquiry may be used to help teach them to give scorable responses. This inquiry may be used liberally with the first two cards but not thereafter, which means that responses on cards 3 through 16 may at times have limited scorable data. The possibility of limited scorable data is the price to be paid for maintaining careful standardization of the RATC procedure. McArthur and Roberts sanction deviations from the standard procedures only in specific instances, such as the examination of severely disturbed children. When standard procedures are not or cannot be followed, they recommend considerable caution in using the scoring procedures.

The scoring system for the RATC consists mainly of coding the content of each story for the presence or absence of 13 profile scales that have implications for personality characteristics. Eight of these are adaptive scales that relate to thematic indications of reliance on others, giving support to others, supporting oneself, limit-setting by authority figures, identification of problem situations, and resolving problems in unrealistic, constructive, or particularly insightful ways. The other five are clinical scales that pertain to thematic manifestations of anxiety, aggression, depression, experiences of rejection, and inability to resolve problems. The test manual provides guidelines and examples that promote reliable scoring of each of these scales.

To facilitate interpretation of the data, the raw scores for each of the 13 profile dimensions are summed over the 16 cards and then plotted on a normatively scaled profile form to yield a visual representation of the data, much in the manner of an MMPI-A profile. The interpretive yield of the data is further enriched by use of an interpersonal matrix, which consists of a tabular representation of the frequency of convergence between particular scales and the figures identified (e.g., cooccurrence of themes of reliance on others with description of interaction with a maternal figure).

4.15.7.5.2 Psychometric foundations

As with the careful coding of specific variables on the TAT, the development of standardized administration and scoring procedures for the RATC has demonstrated that story telling projective methods can achieve psychometric respectability. McArthur and Roberts (1990) report inter-rater agreement ranging from 0.80 to 0.93 in the scoring of the various dimensions in their system and split-half reliabilities ranging from 0.44 to 0.86 on their profile scales, with half of these scales showing reliability coefficients of 0.73 or higher. The adaptive scales appear to work better in this regard than the clinical scales: only one of the five clinical scales (inability to resolve problems) shows a split-half reliability greater than 0.55, whereas all but one of the eight adaptive scales (reliance on others) shows a split-half reliability above 0.60.

Validity studies conducted with 200 nonpatient and 200 outpatient youngsters aged 6–15 indicate that subjects generally tell stories that fit the expectations for the cards for example, that a card intended to depict parental affection typically elicits stories involving affection. In addition to thus showing content validity, the RATC profiles have been found to distinguish between well-adjusted and clinic youngsters, to be more resistant to efforts at manipulation than self-report measures, and to correlate well with a behavior problem checklist completed by subjects' parents (McArthur & Roberts, 1990; Worchel, Rae, Olson, & Crowley, 1992). Significantly, however, in the comparisons between well-adjusted and clinic youngsters, significant differences were found on all eight of the adaptive scales and on two of the five clinical scales, but not on the clinical scales for anxiety, aggression, and depression.

With respect to its normative database, the RATC manual provides the mean and standard deviation for each of the profile scales for 200 nonpatient youngsters divided by gender and by four age categories (6–7, 8–9, 10–12, and 13–15).

4.15.7.5.3 Clinical utility

Similar to the TAT, CAT, and other story-telling measures, the RATC contributes to personality assessment of young people primarily by casting light on their underlying attitudes and concerns. Although the measure includes specifically designated clinical scales that might be expected to assist in clinical diagnosis as well as dynamic analysis of a youngster's personality functioning, its clinical scales as noted appear less reliable than its adaptive scales and less capable of discriminating between patient and nonpatient populations.

The RATC is especially valuable to clinicians as well as researchers by virtue of its careful standardization and adequate psychometric properties. Aside from facilitating the interpretive process and providing the basis for systematic accumulation of data, these features of the instrument make it particularly attractive to examiners conducting forensic evaluations. Having used the RATC in their projective assessment of a child or adolescent, rather than the TAT or CAT, psychologists will be better prepared to justify their procedures and conclusions when giving testimony as an expert witness.

However, examiners may sometimes question whether the information they get from the RATC warrants the amount of time required to administer it, which in our experience approximates an hour on the average. For some young children, moreover, there may be difficulties in sustaining their investment in the task. There is no short form of the RATC to use and no shortcut, such as selecting just a few cards to use or discontinuing a protracted administration before giving all 16 cards. Doing so eliminates the standardization of the instrument and prevents any meaningful scoring or comparisons with normative data, which means that the test loses its advantage over a TAT or CAT interpreted by inspection. Our view in this matter is that the RATC warrants the time required to administer and score it in the standardized manner and, more generally, that psychologists should if at all possible take whatever time is necessary to use standardized instruments when they are available.

The alternate form of the RATC involving African-American figures may prove useful in assessing African-American young people. However, there are as yet no data establishing the reliability and validity of this alternate form, nor have any representative normative data been published for it. Hence the RATC should be used cautiously in multicultural assessments, for which at present the most promising instrument is Tell-me-a-story.

4.15.7.6 Tell-me-a-story

The Tell-me-a-story (TEMAS) was designed as a multicultural story-telling test based on the concept that personality development occurs within a sociocultural system in which individuals internalize the cultural values of their family and society (Costantino, Malgady, & Rogler, 1988). The TEMAS is intended for use with African-American, Hispanic, and Caucasian children and adolescents aged 5–18 and comes in two parallel sets, one for minority and the other for nonminority youngsters. The two sets feature either predominantly Hispanic and African-American characters or predominantly nonminority characters all shown in urban environments. Each set comprises 23 cards, 11 of which have alternate versions for males and females and one of which has alternate versions for children and adolescents.

As a distinctive feature of the TEMAS, many of the pictures portray a split scene showing contrasting or conflicting intrapersonal and interpersonal situations that require some resolution, much in the manner of Kohlberg's (1976) moral dilemma stories. For example, one side of a scene may depict apparent delay of gratification, and the other side an inability to delay gratification. How subjects resolve this conflict in their stories speaks to the adaptiveness of their personality functioning and their stage of moral development.

The TEMAS provides quantitative scales for measuring the adequacy of subjects' adaptation with respect to nine aspects of personality functioning: interpersonal relations, aggression, anxiety/depression, achievement motivation, delay of gratification, self-concept, sexual identity, moral judgment, and reality testing. There are also quantitative scales for four cognitive functions related to how individual process information (reaction time, total time, fluency as reflected in the number of words used

in a story, and omissions of relevant visual details) and for four affective functions as indicated by mood states attributed to the main characters in a story (happy, sad, angry, and fearful). These quantitative scales are supplemented by several qualitative indicators used to describe various other characteristics of the stories.

As stressed by Costantino, Malgady, and Rogler (1988), the TEMAS was developed to overcome limitations of traditional thematic apperception tests and differs from the TAT in several significant ways. These include a focus on interpersonal relationships rather than on intrapsychic dynamics; the use of personally relevant and culturally sensitive stimuli that emphasize meaning rather than ambiguity; the representation of both positive and negative poles of emotions, cognitions, and interpersonal functions, as opposed to the heavy weighting of the TAT stimuli with representations of depression, gloom, anger, and hostility; and the introduction of the joint depiction of contrasting circumstances to elicit expressions of conflict resolution and moral judgment.

4.15.7.6.1 Administration and scoring

Subjects are administered either the minority or nonminority form of the TEMAS, and examiners can choose between a long form, which comprises all 23 cards and requires approximately two hours to administer, and a standard short form, which consists of nine cards and requires one hour to give. In keeping with generally recommended practice in multicultural assessment (Dana, 1993, chap. 6, 1996), young people should be tested in their primary language, and those who are bilingual should be tested by a similarly bilingual examiner. Subjects are instructed to tell a complete story about each picture that indicates what is happening now, what happened before, and what will happen in the future. These instructions may be repeated as often as necessary and supplemented with structured inquiries to elicit information concerning who and where the characters are and what they are thinking and feeling.

The TEMAS is scored by rating stories for the presence of the various cognitive and affective functions and on a four-point scale for the level of adjustment reflected in thematic indications of the personality functions. The resulting scores are then totaled and translated into normalized T scores that are graphed to provide a readily interpretable visual profile. The TEMAS manual provides detailed guidelines and case examples to illustrate these scoring procedures.

4.15.7.6.2 Psychometric foundations

As described in the test manual (Costantino, Malgady, & Rogler, 1988), the TEMAS was carefully developed and standardized over several years prior to its publication. Unfortunately, except for continued work by the test's authors, much of it in the form of paper presentations, there has been little published research concerning the psychometric adequacy of the instrument. On balance, however, the data provided in the test manual appear to indicate a promising beginning in demonstrating its soundness.

With respect to inter-rater agreement, studies with Hispanic, African-American, and Caucasian subjects have indicated that both the minority and nonminority versions of TEMAS can be scored reliably, with agreements between trained examiners generally ranging from 75% to 95% across the various scales. Regarding issues of reliability, on the long form of the TEMAS 11 of the 17 personality, cognitive, and affective scales showed internal consistency (alpha) correlations of 0.74 or higher in the standardization data, and the median for all 17 was 0.83. Internal consistency was lower on the short form, with a median of 0.68. The internal consistency data for the short form must be considered preliminary, however, because in this analysis the short form scores were extracted from the protocols of subjects who had in fact completed the long form, rather than from actual administration of the short form. In an assessment of its retest reliability, the short form was administered twice to 51 behavior problem children over an 18 week interval. Very little stability over time was demonstrated in this study. Costantino, Malgady, and Rogler (1988) suggest several plausible explanations for this disappointing result, such as the narrow range of scores among the subjects. Nevertheless, test–retest reliability remains to be demonstrated for both the short and long forms of the instrument.

Turning to the validity of the TEMAS, ratings by psychologists of the types of personality functions pulled by the stimulus pictures have indicated good agreement with the intent in designing them, thus attesting that the test measures what it purports to measure. In terms of its criterion validity, the TEMAS has been found in several studies to discriminate between patient and nonpatient youngsters in Hispanic and African-American as well as Caucasian samples, and to accomplish this distinction in inner city as well as middle-class settings (Costantino, Malgady, Bailey, & Colon-Malgady, 1989; Costantino, Malgady, Colon-Malgady, & Bailey, 1992; Costantino, Malgady, Rogler, & Tsui, 1988). The TEMAS has not demonstrated any capacity to differentiate specific kinds of disorders, but there is some evidence to suggest that certain story characteristics, including the omission of main characters or events and failure to notice a conflict in the picture (due to lack of attention), may be sensitive to the presence of attention deficit hyperactivity disorder (Costantino, Colon-Malgady, Malgady, & Perez, 1991).

Initial studies reported in the TEMAS manual indicate further that many of its scales correlate significantly with behavior ratings by subjects' mothers and teachers and with behavioral observations of their inclinations toward aggressive or disruptive behavior and their capacities for self-confidence and delay of gratification. Preliminary data indicate further that TEMAS profiles of young people prior to their entering therapy can significantly predict aspects of their treatment outcome.

The TEMAS standardization sample comprised 281 male and 361 female youngsters from the New York City area who ranged in age from 5 to 13 years. This sample of 642 children and adolescents included groups of Caucasian, African-American, Puerto Rican, and other Hispanic subjects from predominantly lower- and middle-income families. The test manual lists the mean and standard deviation for each scale by gender and ethnicity for the age groups 5–7, 8–10, and 11–13.

4.15.7.6.3 Clinical utility

The TEMAS brings to clinical assessments the distinct advantages of a well conceptualized and quantitatively standardized story-telling technique that is also culturally sensitive and normed for minority groups of young people. As in the case of the RATC, its coded scales and profile graphs facilitate interpretation and provide the type of documentation that typically proves valuable in forensic cases. Moreover, this measure stands alone as a story-telling test proved applicable in the assessment of African-American and Hispanic children and adolescents. The TEMAS accordingly merits serious consideration for inclusion in a test battery for evaluating personality functioning in young people, especially if they come from an urban minority background and if there are forensic issues in the case.

At this stage in its development, however, the TEMAS has some drawbacks that examiners should keep in mind. First, although it appears to be useful for assessing level of adjustment and incorporation of societal norms among both minority and nonminority youth, TEMAS does not provide the breadth of information concerning specific types of concerns and relationships that emerge from TAT, CAT, and RATC analyses. Second, the normative data available thus far are limited to 5–13-year-old city dwellers and thus do not provide a psychometric basis for drawing conclusions about adolescents aged 14–18, or suburban and rural dwelling youngsters. Third, the length of time (two hours) required to administer the 23 card long-form of the test is often impractical in clinical evaluations. The nine card short-form is an attractive alternative but, as noted, adequate reliability has not yet been demonstrated for the short-form.

4.15.7.7 Draw-a-person

The use of human figure drawings as a projective method of personality assessment is based on the expectation that how subjects draw people will reveal aspects of how they perceive themselves and feel about others. Clinical use of drawing techniques originated in work with children, and the first formal drawing test, called the draw-a-man, was developed by Goodenough (1926) and later refined by Harris (1963) and Naglieri (1988) for use as a nonverbal measure of intellectual development. Machover (1948) introduced the notion of using human figure drawings as a projective device that generates nonverbal, symbolic messages concerning subjects' impulses, anxieties, and conflicts, and she proposed numerous possible meanings for structural features of drawings (e.g., where figures are placed on the page) and the manner in which various parts of the body are drawn (e.g., a disproportionaly large head).

Koppitz (1968, 1984) subsequently focused attention specifically on draw-a-person (DAP) assessment of young people and used on Machover's interpretive hypotheses to formulate 30 specific indicators of emotional disturbance involving the quality of drawings (e.g., asymmetry, transparencies), special features (e.g., teeth showing, arms clinging to body), and omission of body parts (e.g., no eyes, no arms). Naglieri and his colleagues (Naglieri, McNeish, & Bardos, 1991; Naglieri & Pfeiffer, 1992) have developed further coding refinements to produce the draw-a-person: screening procedure of emotional disturbance (DAP:SPED). The DAP:SPED is an actuarialy derived and normatively based system comprising 55 objectively scorable items, such as the measured dimensions and placement of figures. It is intended as a screening test for classifying young people aged 6–17 with respect to their

likelihood of having adjustment difficulties that call for further evaluation.

4.15.7.7.1 Administration and scoring

The DAP is administered by giving subjects a plain 8.5 × 11 inch piece of paper and asking them to draw a person. When they have finished the drawing, they are given another piece of paper and asked to draw a person of the opposite sex from the one they have just drawn. Subjects are further instructed to draw the figure of a whole person rather than a cartoon or stick figure. In keeping with further suggestions by Machover (1951), subjects are also typically asked and to draw a picture of themselves and to provide some thematic content concerning the figures they have drawn.

As elaborated by Handler (1985), there are three alternative ways of eliciting such thematic content: by asking subjects to associate to their drawings, by asking them to make up a story about the people they have drawn, or by asking them specific questions about their drawings. Machover (1951) provided 31 specific questions to be used for this purpose with children, such as "What is their ambition?," "How happy are they?," "What do they worry about?," and "What are their good points?"

As with story-telling techniques, figure drawing methods are most commonly interpreted in clinical practice by an inspection in which personality characteristics are inferred primarily from subjective impressions of noteworthy or unusual features of the figures drawn. Items included in the Koppitz and DAP:SPED scales may often enter into these impressionistic assessments. However, these scales serve only to identify emotional disturbance without contributing in other ways to personality description, such as by indicating how individuals process information, handle emotion, and manage stress. They have consequently not been widely adopted clinically, and there are no other DAP scoring systems that have attracted much attention in the research literature.

4.15.7.7.2 Psychometric foundations

There is very little psychometric foundation for traditional applications of the DAP, and clinical use of this instrument in assessing young people frequently goes well beyond any empirical justification in empirical data. The influential inspection approach used by Machover is neither standardized nor codified, which precludes any systematic evaluation of its reliability or the accumulation of a normative database. Additionally, although it may be reasonable to expect that unusual emphasis on or omission of some body part will reflect some particular concern about the nature or functions of that body part, most of Machover's specific hypotheses concerning the symbolic significance of figure drawing characteristics lack consistent research support (Kahill, 1984; Roback, 1968; Swensen, 1957, 1968).

The Koppitz system is adequately codified and intended to comprise items that have demonstrably low occurrence in the normal population. However, there is no normative database for the system, its reliability is yet to be demonstrated, and there is some question as to whether it can differentiate between well-adjusted and emotionally disturbed children. In a carefully done study in which Tharinger and Stark (1990) compared groups of mood disordered, anxiety disordered, mood/anxiety disordered, and well-adjusted youngsters aged 9.5–14.75 years, the Koppitz signs showed good interscorer agreement but did not differentiate among the subject groups either in mean total score or in the frequency of any of the 30 individual items.

Such findings warrant concern that the DAP may be a faulty projective technique with questionable propriety for continued clinical use. However, it could be that the DAP is a potentially sound method for which there has not yet been sufficiently sophisticated development and evaluation to document its capacities. The previously mentioned work of Naglieri et al. (1991) on the DAP:SPED appears to speak to this point. The DAP:SPED was standardized on a representative national sample of 2355 children and adolescents aged 6–17 years; its objective scoring procedures have generated inter-rater agreements above 90%; its internal consistency (alpha) reliability estimates are 0.76 among 6–8-year-olds, 0.77 among 9–12-year-olds, and 0.71 among 13–17-year-olds; and its normalized total score has shown substantial capacity to differentiate nonpatient youngsters from those with identified behavioral or emotional problems (McNeish & Naglieri, 1991; Naglieri & Pfeiffer, 1992).

Also noteworthy is the work of Tharinger and Stark (1990), who paired their failure to validate the Koppitz system with an investigation of a proposed new system of their own, the DAP integrative system. The integrative DAP is based on a qualitative holistic scoring approach in which drawings are given an overall adjustment rating on a scale from 1 (absence of psychopathology) to 5 (severe psychopathology). Examiners are instructed to base their impressions on their integrated sense of four characteristics of a drawing, with the pathological end of the scale involving (i) inhumanness

of the drawing suggesting feelings of being incomplete, grotesque, or monstrous; (ii) lack of agency as conveyed by a sense of powerlessness; (iii) lack of well-being as reflected in negative facial expressions; and (iv) a hollow, vacant, or stilted sense indicating lack of capacity to interact.

In the same study in which Tharinger and Stark failed to validate the Koppitz system, their integrative system total score significantly discriminated the mood disordered and anxiety/mood disordered subjects from the well-adjusted youngsters and also correlated significantly with the Coopersmith self-esteem inventory, thus attesting the capacity of the DAP to depict a youngster's sense of self. Contemporary literature abounds with sharply divided opinion concerning whether the DAP is a worthless test that should no longer be used (e.g., Gresham, 1993; Motta, Little, & Tobin, 1993) or is instead a potentially valuable clinical tool that has too often been carelessly used or inadequately researched (e.g., Bardos, 1993; Holtzman, 1993). Results to date with the DAP:SPED and the DAP integrative system give some reason to believe that improved methodology may yet establish sound psychometric foundations for carefully specified applications of the DAP.

4.15.7.7.3 Clinical utility

Despite its widespread use in describing subjects' personality characteristics, there is virtually no empirical evidence that traditional DAP interpretation has any clinical utility. Smith and Dumont (1995) asked a group of experienced psychologists and graduate students who had been trained in the use of the DAP to review and comment on a case file. They found that these clinicians routinely utilized specific symbolic representations in the drawings to draw inferences about the client's personality characteristics and diagnostic status—even though research does not support any such isomorphic correspondence of specific drawing characteristics to specific features of personality. Aside from the general ethical issues of using test instruments in unwarranted ways, examiners who are preparing forensic testimony jeopardize their credibility by employing the DAP in this manner.

This does not mean that human figure drawings are without utility in the clinical assessment of children and adolescents. First, by contrast with the practical difficulties of administering the full RATC and TEMAS, the DAP is an easily administered measure that requires no test stimuli and usually takes less than 10 minutes.

Second, as a nonverbal measure the DAP can prove especially useful in the evaluation of frightened, reticent, or otherwise uncommunicative young people, and it is unaffected by language difficulties or bilingualism. As noted by Cummings (1986), moreover, many young children may be more capable of expressing their thoughts and feelings in drawings than in words, and drawing pictures is a more familiar activity to most youngsters than most of the tasks that are set for them in a psychological examination.

Third, unusual characteristics of drawings can suggest avenues for further exploration, even in the absence of definitive conclusions about what these features signify. Examiners need to be circumspect in pursuing such avenues, however, lest they prematurely conclude that they are exploring in correct directions. The pursuit of speculative hypotheses is just as likely to lead into blind alleys as down fruitful paths.

Finally, there is ample indication in the contemporary literature that such approaches as the DAP:SPED and the DAP Integrative System can help to identify the presence and severity of emotional disturbance in general and can accordingly contribute to treatment recommendations, planning, and monitoring. In the future, codifiable and interpretable thematic content in subjects' stories about their drawings, comparable to the data derived from story telling techniques, may further enhance the utility of the DAP.

4.15.7.8 House-tree-person

The House-tree-person (HTP) test was devised by Buck (1948, 1985) as a means of tapping the concerns, interpersonal attitudes, and self-perceptions of young people more fully than is possible with human figure drawings alone. The HTP is intended for use with anyone over the age of three and was regarded by Buck as a nonthreatening instrument that can serve well to minimize a child's anxiety in testing situations and to assess personality functioning in multicultural and bilingual settings. As postulated by Buck and subsequently elaborated by Hammer (1958, 1985), subjects' drawings of these three objects are considered to provide symbolic representations of important aspects of their world.

Specifically, the house, as a dwelling place, is expected to arouse feelings toward the subject's home life and family relationships and, particularly for children, attitudes toward their parents and siblings. The tree is seen as encouraging projection of personal feelings

about the self that would be more anxiety-provoking to express in drawing a person, because the latter is more obvious in its representation as a self-portrait. More specifically, the way the trunk of the tree is drawn is considered to portray a subject's feeling of basic power and inner strength; the branches are seen as depicting the subject's ability to derive satisfaction from the environment; and the overall organization of the drawing is taken as a reflection of the individual's feeling of intrapersonal balance. Finally, the drawing of the person is expected to reveal aspects of how subjects view themselves, how they would like to be, and what they think about significant other people in their lives.

4.15.7.8.1 Administration and scoring

Although Buck recommends a four-page booklet with pages measuring 7 × 8.5 inches, most HTP examiners use four sheets of standard 8.5 × 11 in paper on which subjects are asked first to draw "as good a picture of a house as you can," then a tree, and then a person of each sex. Subjects are told they can take us long as they wish and draw any kind of house, tree, or person they like. Completion of the drawings is followed by an interrogation phase in which numerous questions devised by Buck are used to encourage subjects to define, describe, and associate to their drawings (e.g., "About how old is that tree?," "What does that house make you think of?" "Is that person happy?" Buck also recommended a chromatic phase of the HTP in which subjects would do a second rendering of their drawings in crayon rather than pencil, followed by another interrogation. There are no data to indicate the frequency with which examiners conduct Buck's full HTP administration or instead limit the test to the pencil drawings, without either an inquiry or a chromatic phase.

Following the example of Goodenough, Buck originally proposed an elaborate quantitative system for objective coding of structural features of the HTP drawings, such as their size and proportions. As best as can be determined from the literature, however, quantitative coding of the HTP has rarely been employed in clinical practice. Instead, clinicians using this instrument typically rely on a qualitative inspection technique to identify symbolic implications of drawing characteristics for aspects of personality functioning. As in the case of interpreting the DAP in the tradition of Machover, many of the interpretive hypotheses for the HTP suggested by Buck and by Hammer are quite specific. A door that is tiny in relation to the size of the house is interpreted as indicating reluctance to make contact with the environment and an inhibited capacity for social relations, for example, and overemphasis on the roots of the tree where they make contact with the ground is taken as evidence of subjects' concerns about losing their grip on reality.

4.15.7.8.2 Psychometric foundations

The entire psychological literature contains only a handful of articles bearing on the psychometric foundations of the HTP. Most of these are over 25 years old, and none of them provides convincing supportive evidence for the interpretive uses of the instrument recommended by its leading proponents. Buck's (1985) 350-page revised manual contains extensive guidelines and case illustrations to facilitate interpretation, but neither reliability nor validity appears in the index. As intriguing as the rationale for the instrument may be to clinicians, especially those who are psychodynamically oriented, there is at present no empirical basis to warrant inferring personality characteristics from it.

4.15.7.8.3 Clinical utility

Similar to the DAP, the HTP offers the potential advantages in clinical practice of a brief, easily administered, nonverbal, and largely culture-free assessment instrument, along with perhaps being even less anxiety provoking than the DAP. Emerging refinements with the DAP suggest that the HTP as well might prove useful in identifying maladjustment in general, monitoring progress and change in psychotherapy, and even in pointing to possible specific areas of conflict and concern. Moreover, like the DAP inquiry, the HTP interrogation can produce story-telling content that in turn can be codified and suggest topics for further exploration. As matters presently stand, however, any such utility remains an unfulfilled potential, and examiners should be circumspect about including the HTP in their test batteries and basing any firm conclusions on it.

4.15.7.9 Kinetic Family Drawing

Machover (1948) and numerous other clinicians who pioneered in using the DAP to assess young people suggested that useful information might also be obtained by asking individuals to draw members of their family. This suggestion was formalized by Burns and Kaufman (1970, 1972) as the kinetic family drawing (KFD) technique, in which subjects are instructed to draw a picture of everyone in their family, including themselves, doing something. These drawings are then examined for such objective features as omissions of body parts or of members of the family and interpreted according to the actions, styles, and symbols represented in them.

Actions in the Burns and Kaufman approach refer to the ways in which the figures drawn are behaving toward each other, which are thought to provide clues to the intensity and emotional tone of their relationships. Styles concern barriers between family members that prevent them from interacting at all, which may be expressed by drawing some of them at a far distance from the others or encasing them in a circle or a box. Symbols comprise a list of specific items, such as beds, flowers, stoves, cats, and the like, the inclusion of which is considered to reflect various specific unconscious impulses or concerns.

Publication of the KFD was followed by a school adaptation of the technique by Prout and Phillips (1974), known as the kinetic school drawing (KSD), in which children are asked to draw a school picture of themselves, their teacher, and a friend or two in which everyone is doing something. The KSD was intended to provide information about peer relationships and about attitudes and concerns related to school in the same manner as the KFD does for family relationships and feelings about the home.

Knoff and Prout (1985a, 1985b) subsequently recommended combining the KFD and KSD and administering both measures for purposes of analysis and comparison. This combined approach, which they call the kinetic drawing system, is expected to identify adjustment difficulties both at home and in school, to clarify causal or reciprocal relationships between family and school-related issues, and to indicate which people in subjects' lives (e.g., father, sister, teacher) are sources of support of tension. In a further extension of this approach, Burns (1987) has developed a kinetic house-tree-person, in which subjects are instructed to draw on a single page a picture of a house, a tree, and a person in "some kind of action."

4.15.7.9.1 Administration and scoring

Similar to procedures followed in other drawing techniques, administration of the KFD consists of giving subjects an 8.5 × 11 inch piece of plain paper and asking them to draw a picture of everyone in their family, including themselves, doing something. Beginning with an interpretive approach modeled basically after Machover's qualitative system, Burns (1982) developed a long list of actions, style, and symbol characteristics for examiners to consider. As summarized by Knoff and Prout (1985a, 1985b), at least four objective methods for coding characteristics of kinetic drawings have also been proposed by various investigators, but none of these has become consistently visible in the literature. A review of the KFD literature by Handler and Habenicht (1994) indicates in general that cumulative knowledge concerning the adequacy and utility of this instrument has been limited by considerable variation in whether and how it has been scored in research studies and applied in clinical practice.

Also of note is a thematic elaboration of the KFD proposed by McConaughy and Achenbach (1994) and called the semistructured interview protocol. In this approach subjects are given the following question to answer after they have completed their family drawing: (i) what are they doing, (ii) what kind of person is [each member], (iii) what are three words that describe [each member], (iv) how does [each member] feel in this picture, (v) what is [each member] thinking, (vi) who do you get along with best, (vii) who do you get along with least, and (viii) what is going to happen next in your picture?

4.15.7.9.2 Psychometric foundations

Handler and Habenicht (1994) were able to reference a substantial number of publications concerning the KFD. However, they were forced to conclude that, despite its widespread use, this projective method has not yet been adequately developed with respect to its psychometric properties. Most of the systems previously proposed for coding KFD characteristics can achieve substantial inter-rater reliability, with percentages of agreement generally ranging well above 0.85 in various studies (Cummings, 1986; Handler & Habenicht, 1994). However, none of these scoring systems has demonstrated satisfactory retest reliability, even after very brief intervals, and there is little empirical basis for challenging the opinion that "the KFD still remains primarily a clinical instrument with inadequate norms and questionable validity" (Handler & Habenicht, 1994, p. 441).

With this in mind, Handler and Habenicht have recommended a holistic, integrative approach to KFD interpretation, much in the manner of the previously noted integrative DAP method used by Tharinger and Stark (1990), rather than the coding and summation of lists of signs and symbols. In the same study in which Tharinger and Stark demonstrated the superiority of their integrative DAP to the Koppitz scoring system, they also found that a holistic

method of evaluating kinetic drawings discriminated adjusted from maladjusted children more effectively than a traditional scoring guide for KFD interpretation developed by Reynolds (1978).

In developing his scoring guide, Reynolds urged clinicians and researchers to avoid cookbook approaches that fail to go beyond positing direct and absolute links between individual drawing characteristics and specific personality features. Instead, scores for drawing characteristics should be only a first step in interpreting drawings as gestalts that take on meaning only in relation to a youngster's personal, family, and cultural context. It can be hoped that eventual attention to this sound advice, perhaps through the standardization of integrated coding systems, will establish for kinetic drawings the validity that is presently lacking for them to be used with confidence in clinical evaluations.

4.15.7.9.3 Clinical utility

Conceived as a means of understanding how children and adolescents conceptualize their family and perceive themselves within the family context, the KFD is a potentially useful instrument for elaborating the interpersonal dynamics of young people and orchestrating individual or family therapy for those with adjustment difficulties. By employing the kinetic drawing system, examiners can explore school-based as well as family-related issues for such purposes. Of further potential benefit, subjects from a variety of ethnic and minority group backgrounds have been found to express dimensions of their family culture in their KFDs (Handler & Habenicht, 1994), which suggests a role for this instrument in multicultural assessment.

Unfortunately, however, the previously noted psychometric limitations of the KFD makes its clinical use problematic at present, particularly with respect to basing any firm conclusions on what and how subjects draw. Until such time as adequate research has clarified what the KFD and other projective drawings methods can and cannot do, clinicians are well advised to regard their figure drawing findings as suggesting but not confirming any notions about the subject. We would endorse in this regard the conclusion of Knoff (1990) that "The hypothesis generating use of projective drawings and their ability to be interpreted within various psychological orientations remains both viable and defensible [but] the validity of hypotheses tied to specific drawing characteristics must still be determined" (p. 99).

4.15.7.10 Sentence Completion Methods

Sentence completion methods consist of initial words or phrases, called stems, that subjects are asked to extend either orally or in writing into complete sentences. Typically used stems vary in length and structure from just one or two words, such as "People ..." or "I wish ...," to detailed specification of people or situations, such as "If only my mother would ..." or "When he found he had failed the examination, he ..." As in the case of responses on other projective methods, the manner in which subjects complete sentences is expected to provide an indirect source of information concerning their underlying feelings, attitudes, and level of adjustment. In addition to eliciting general information in these respects, the specificity in many sentence completion stems encourages subjects to reveal their orientation toward particular events and circumstances in their lives.

Sentence completion methods originated in word association tests dating back to the 1890s, and, as reviewed by Haak (1990) and Lah (1989b), were developed during the 1940s and 1950s into a large number of formal and informal versions, of which the most noteworthy published scales were the Rohde, the Sacks, the Forer, and the Miale-Holsopple sentence completion tests and the Rotter incomplete sentences blank (RISB). Of these, the RISB has become the best known and most widely used and includes an adult, a college, and a high school form (Rotter, Lah, & Rafferty, 1992). Also available are more recently developed sentence completion forms constructed by Brown and Unger (1992) for use with adolescents and adults and by Hart (1986) for evaluating school age children. Further comments focus primarily on the RISB, as the predominant and most recently revised exemplar of the sentence completion method, and on the Hart sentence completion test (HSCT), because of its specificity for assessing young people.

4.15.7.10.1 Administration and scoring

The RISB is a 40-item test printed on the front and back of a one-page form and comprised mainly of brief stems (eight have just a single word, 21 have two words, six have three words, and the remaining five have four words). Subjects are given a pencil and asked to "complete these sentences to express your real feelings." The test can be administered either individually or in a group setting; however, Rotter et al. (1992) caution against taking oral rather than written responses, primarily because doing so injects an interpersonal component into the testing administration that can influence subjects' responses in various confounding ways.

The RISB and other sentence completion tests are typically interpreted in clinical practice by the inspection methods we have described previously; that is, examiners read the content of the items and from impressions of what they signify concerning a subject's probable personality characteristics. Rotter et al. (1992) also provide a scoring system for rating each item on a seven-point scale from 0 (most positive adjustment) to 6 (most indication of conflict). These ratings are totaled to yield an overall adjustment score. There is little indication that this or any other codification has received much attention in clinical practice or research studies, even though some work with college students indicated that the overall RISB score can discriminate those who are receiving counseling or psychotherapy from their nonpatient peers (Lah, 1989a).

The HSCT is also a 40-item measure in which the item stems were designed specifically to use with children and to sample family, social, school, and self dimensions of their lives (Hart, Kehle, & Davies, 1983). HSCT scoring involves rating each item as negative, neutral, or positive with respect to adjustment and also rating 10 scales composed of various clusters of items concerned with perceptions of self, family, and school on a five-point negative-to-positive continuum. Criteria are provided for each item to guide examiners in their ratings and thereby to enhance scoring objectivity and the prospects for achieving inter-rater agreement.

4.15.7.10.2 Psychometric foundations

Although there is a substantial RISB research literature, most of the published studies have used the instrument as a measure of adjustment but, as pointed out by Lah (1989a), they were not designed to evaluate its properties. Particularly with respect to its validity, there has been virtually no accumulation of empirical evidence to support any diagnostic or predictive inferences about personality functioning beyond the modestly demonstrated capacity of the RISB to identify maladjustment. Likewise, little systematic progress has been made in assessing the reliability of RISB data and establishing normative standards for them.

This is a regrettable state of affairs because, aside from being a psychodynamically compelling method of enriching personality assessment, the RISB is an eminently codeable instrument for which excellent interscorer agreement has been demonstrated using the Rotter et al. approach (Lah, 1989a). Moreover, as discussed in the introductory portions of this chapter, sentence completion responses have numerous objective, structural, and behavioral components (e.g., time to completion, length of sentences, frequency of self- vs. other-reference, items omitted) that have rarely been considered in codifying these methods or attempting to establish a psychometric foundation for them.

The HSCT coding system has also achieved substantial inter-rater agreement and some preliminary success differentiating between emotionally disturbed and well-adjusted children. However, Hart (1986, p. 269) observed that it remained to further research to demonstrate adequate reliability and criterion validity for his instrument and to develop sufficiently broad normative data. The available literature, circa 1996, suggests that not much if any, progress has been made in this regard.

4.15.7.10.3 Clinical utility

Sentence completion methods bring to the personality assessment battery an easily administered projective test that can be given to groups as well as individuals and requires only about 30 minutes on average for subjects to complete. The scorable 40-item RISB high school form is suitable for high school and most middle school youngsters, and the 40-item HSCT can be used comfortably with younger children. Barring a language or reading difficulty, the meaning of the sentence completion stems is clear, and subjects are not asked to explain, elaborate, or otherwise account for their responses. Hence, incomplete sentence methods are often less threatening or anxiety-provoking than inkblot, story-telling, and figure drawing procedures. However, the previously noted relative lack of ambiguity in sentence completion stems means that subjects are more aware than on other projective methods of how they are presenting themselves through their responses.

At this point in time, the RISB and HSCT can be used with justification to form general impressions of a young person's level of adjustment and to formulate hypotheses concerning possible conflicts or concerns the young person has and how he or she may feel about self, other people, and certain situations. Such hypotheses can justifiably be expressed only as speculations, however, not as conclusions, and they should be considered reasonably correct only to the extent that they are supported by other reliable sources of information.

Haak (1990) has asserted that sentence completion methods can be used effectively

with children to rule out intellectual difficulties, attention deficit disorder, and stress and to rule in depression, anxiety, thought disturbance, and defensiveness, and provides detailed clinical guidelines for doing so. These suggestions seem sensible in each case, but they are entirely unverified by empirical data. Hence, like other impressions formed by experienced clinicians, Haak's diagnostic guidelines provide an agenda for confirmatory research, but they should not be elevated from speculations to the status of conclusions until that empirical confirmation becomes available.

4.15.8 FUTURE DIRECTIONS

Projective methods have been extensively used, taught, and studied for many years, and it seems likely that they will continue in the future to be regularly included in test batteries for assessing personality functioning in young people. Survey data cited in this chapter document that over the last 30 years psychologists have continued with much the same frequency to apply these instruments in practice and utilize them in research. During this same period the psychometric adequacy and clinical contributions of projective methods have been regularly and vigorously challenged, and this chapter indicates that the uses to which some of these instruments are put often go beyond available justification in empirical data.

How do we account for the persistent use by clinicians of many such as yet unvalidated methods? One could say that there just happen to be legions of uninformed or unethical clinicians in practice who do not hesitate to employ useless methods if doing so serves their purpose in some way. However, it seems doubtful that a vast segment of the profession deserves to be tarred with such a broad brush of evil. Rather, at least the majority of professionals who use projective methods must have good reason on the basis of their clinical and individual case experience to believe sincerely that these methods provide valid and useful information about personality functioning in ways that have not yet been translated into supporting research data. One could then say that such clinical confidence in projective methods is based solely on illusory correlation. However, it seems doubtful that a vast segment of professional personality assessors could be so thoroughly deluded.

These observations suggest that projective techniques have been and will continue to be used because they yield valuable information and generate fruitful hypotheses in clinical assessments. Hopefully, adequate research methods will lead the way toward improved standardization and codification of projective methods, that will in turn enhance the psychometric foundations on which they rest. As Kuehnle (1996) has pointed out, inappropriate use of projective instruments for purposes for which they have not been validated, such as identifying children as having been traumatized or sexually abused, violates ethical standards and risks causing harm to young people and their families.

More than any of the other projective methods, the Rorschach has had the benefit of rigorous attention to research methodology, as demonstrated in publications by Exner (1995) and Weiner (1995a) and, as indicated earlier in this chapter, this method presently rests on a solid psychometric foundation. What the future of projective methods needs is similar methodological attention to documenting the reliability of other techniques and their validation for various purposes. To say that sophisticated research methods are not applicable to projective methods is as unwarranted as asserting that these methods are by nature invalid. To say that projective test responses cannot be examined scientifically without detracting from their idiographic richness sells projective methods short and prevents them from realizing their full potential.

Along with these needs for improved research and a narrowed gap between data and practice, the most important future direction for projective testing of young people lies in developing adequately representative normative data that will facilitate age-specific and multicultural assessment. There is no lack of projective techniques, but there is a decided lack of reliable information concerning how children and adolescents of different ages, from diverse backgrounds, and with various kinds of personality strengths and weaknesses, should be expected to respond to them. In addition to the nearly empty coffers of cross-sectional normative data of these kinds, the cupboard is virtually bare with respect to longitudinal ata indicating how young people are likely to change over time and with maturation in how they respond to projective tests. Collecting and disseminating such data is the number one agenda item for the future development of projective assessment of children and adults.

4.15.9 SUMMARY

Projective tests are methods of personality assessment in which some degree of ambiguity in the test stimuli or instructions creates opportunities for subjects to structure their responses in terms of their individual personality characteristics, and thereby provide information about the nature of these characteristics. Although projective methods are accordingly more ambiguous and less structured than so-called objective methods, the differences between these methods are relative rather than absolute. All projective tests contain objective as well as subjective features and elicit responses that are representative as well as symbolic of behavior, and they differ from each other in the extent to which they are ambiguous.

Because of their relatively unstructured nature, projective tests measure personality functioning in subtle and indirect ways and tap underlying psychological characteristics at a less conscious level than relatively structured measures. Projective test data consequently provide valuable information about how people are likely to think, feel, and act that is difficult to obtain from objective assessment procedures, and they are also less susceptible than objective test data to the influence of test-taking attitudes.

Projective methods can be used to good effect with children and adolescents as well as adults. The basic interpretive conclusions and hypothesis that attach to projective test variables apply regardless of the age of the subject, provided that examiners determine the implications of their data in the light of normative developmental expectations. Surveys of clinical and school settings indicate that the projective instruments most frequently administered in evaluating young people are the Rorschach inkblot method, the thematic apperception test (TAT), the children's apperception test, the Roberts apperception test for children (RATC), the tell-me-a-story (TEMAS) test, the draw-a-person, the house-tree-person, the kinetic family drawing, and alternate forms of the sentence completion test.

For each of these nine projective tests, this chapter reviews their composition, administration, scoring, psychometric foundations, and clinical utility. Despite widespread utilization of these nine tests in clinical practice to draw conclusions about the personality characteristics, level of adjustment, and treatment needs of young people, only the Rorschach presently rests on a solid empirical foundation. Properly collected and interpreted, Rorschach data provide numerous demonstrably reliable and valid indices that facilitate differential diagnosis and treatment planning for children and adolescents with adjustment difficulties.

The RATC and TEMAS have shown through the development of standardized administration and scoring procedures that story-telling methods have the potential to achieve psychometric respectability. However, further research on the RATC and TEMAS is needed with regard to multicultural and adolescent norms, respectively. With regard to the TAT, recently emerging psychometrically sound schemes for coding specific personality characteristics reflected in thematic content, such as SCORS, appear to provide a basis for empirical decision making.

The other measures reviewed have in various ways also shown potential to be codified and refined on the basis of empirical data, and there is reason to be hopeful that advances in research methodology will eventually close a currently regrettable gap between what is known for sure about these projective methods and what is frequently assumed to be true about them in clinical practice. Until this gap is narrowed, and especially until such time as more extensive normative and multicultural data become available, most clinical inferences from projective data should be regarded as hypotheses to be confirmed rather than as facts on which to base conclusions and recommendations.

4.15.10 REFERENCES

Abraham, P. P., Lepisto, B. L., Lewis, M. G., & Schultz, L. (1994). Changes in Rorschach variables of adolescents in residential treatment: An outcome study. *Journal of Personality Assessment, 62*, 505–514.

Alvarado, N. (1994). Empirical validity of the Thematic Apperception Test. *Journal of Personality Assessment, 63*, 59–79.

Ames, L. B., Metraux, R. W., Rodell, J. L., & Walker, R. N. (1974). *Child Rorschach responses* (Rev. ed.). New York: Brunner/Mazel.

Ames, L. B., Metraux, R. W., & Walker, R. N. (1971). *Adolescent Rorschach responses* (Rev. ed.). New York: Brunner/Mazel.

Archer, R. P., Imhof, E. A., Maruish, M., & Piotrowski, C. (1991). Psychological test usage with adolescent clients: 1990 survey findings. *Professional Psychology, 22*, 247–252.

Archer, R. P., & Krishnamurthy, R. (1993a). A review of MMPI and Rorschach interrelationships in adult samples. *Journal of Personality Assessment, 61*, 277–293.

Archer, R. P., & Krishnamurthy, R. (1993b). Combining the Rorschach and MMPI in the assessment of adolescents. *Journal of Personality Assessment, 60*, 132–140.

Atkinson, J. W., & Feather, N. T. (1966). *A theory of achievement motivation*. New York: Wiley.

Bardos, A. N. (1993). Human figure drawings: Abusing the abused. *School Psychology Quarterly, 8*, 177–181.

Bellak, L. (1993). *The T.A.T., C.A.T., and S.A.T. in clinical use* (5th ed.). Boston: Allyn & Bacon.

Bellak, L., & Hurvich, M. (1966). A human modification of the Children's Apperception Test. *Journal of Projective Techniques, 30*, 228–242.

Bellak, L., & Siegel, H. (1989). The Children's Apperception Test (CAT). In C. S. Newmark (Ed.), *Major psychological assessment instruments* (Vol. II, pp. 99–127). Boston: Allyn & Bacon.

Bornstein, R. F. (1995). Sex differences in objective and projective dependency tests: A meta-analytic review. *Assessment, 2*, 319–331.

Buck, J. N. (1948). The H-T-P technique, a qualitative and quantitative method. *Journal of Clinical Psychology, 4*, 317–396.

Buck, J. N. (1985). *The House-Tree-Person technique: Revised manual.* Los Angeles: Western Psychological Services.

Burns, R. C. (1982). *Self-growth in families: Kinetic Family Drawings (K-F-D) research and application.* New York: Brunner/Mazel.

Burns, R. C. (1987). *Kinetic-House-Tree-Person drawings (K-H-T-P).* New York: Brunner/Mazel.

Burns, R. C., & Kaufman, S. H. (1970). *Kinetic Family Drawings (K-F-D): An introduction to understanding children through kinetic drawings.* New York: Brunner/Mazel.

Burns, R. C., & Kaufman, S. H. (1972). *Actions, styles, and symbols in Kinetic Family Drawings (K-F-D).* New York: Brunner/Mazel.

Butcher, J. N., & Rouse, S. V. (1996). Personality: Individual differences and clinical assessment. *Annual Review of Psychology, 47*, 87–111.

Chandler, L. A. (1990). The projective hypothesis and the development of projective techniques for children. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (Vol. 2, pp. 55–69). New York: Guilford Press.

Costantino, G., Colon-Malgady, G., Malgady, R. G., & Perez, A. (1991). Assessment of attention deficit disorder using a thematic apperception technique. *Journal of Personality Assessment, 57*, 97–95.

Costantino, G., Malgady, R. G., Bailey, J., & Colon-Malgady, G. (1989). *Clinical utility of TEMAS: A projective test for children.* Paper presented at the meeting of the Society for Personality Assessment, New York.

Costantino, G., Malgady, R. G., Colon-Malgady, G., & Bailey, J. (1992). Clinical utility of the TEMAS with non minority children. *Journal of Personality Assessment, 59*, 433–438.

Costantino, G., Malgady, R. G., & Rogler, L. H. (1988). *TEMAS (Tell-Me-A-Story) manual.* Los Angeles: Western Psychological Services.

Costantino, G., Malgady, R. G., Rogler, L. H., & Tusi, E. C. (1988). Discriminant analysis of clinical outpatients and public school children by TEMAS: A thematic apperception test for Hispanics and Blacks. *Journal of Personality Assessment, 52*, 670–678.

Cramer, P. (1987). The development of defense mechanisms. *Journal of Personality, 55*, 597–614.

Cramer, P., & Blatt, S. J. (1990). Use of the TAT to measure change in defense mechanisms following intensive psychotherapy. *Journal of Personality Assessment, 54*, 236–251.

Cummings, J. A. (1986). Projective drawings. In H. M. Knoff (Ed.), *The assessment of child and adolescent personality* (pp. 199–204). New York: Guilford Press.

Dana, R. H. (1985). Thematic Apperception Test (TAT). In C. Newmark (Ed.), *Major psychological assessment instruments* (pp. 89–134). Boston: Allyn & Bacon.

Dana, R. H. (1993). *Multicultural assessment perspectives for professional psychology.* Boston: Allyn & Bacon.

Dana, R. H. (1996). Culturally competent assessment practice in the United States. *Journal of Personality Assessment, 66*, 472–487.

Di Leo, J. H. (1983). *Interpreting children's drawings.* New York: Brunner/Mazel.

Elbert, J. C., & Holden, E. W. (1987). Child diagnostic assessment: Current training practices in clinical psychology internships. *Professional Psychology, 18*, 587–596.

Exner, J. E., Jr. (1991). *The Rorschach: A comprehensive system. Vol. 2. Interpretation* (2nd ed.). New York: Wiley.

Exner, J. E., Jr. (1993). *The Rorschach: A comprehensive system. Vol. 1. Basic foundations* (3rd ed.). New York: Wiley.

Exner, J. E., Jr. (Ed.) (1995). *Issues and methods in Rorschach research.* Mahwah, NJ: Erlbaum.

Exner, J. E., Jr., & Andronikoff-Sanglade, A. (1992). Rorschach changes following brief and short-term therapy. *Journal of Personality Assessment, 59*, 59–71.

Exner, J. E., Jr., Thomas, E. A., & Mason, B. (1985). Children's Rorschachs: Description and prediction. *Journal of Personality Assessment, 49*, 13–20.

Exner, J. E., Jr., & Weiner, I. B. (1995). *The Rorschach: A comprehensive system. Vol. 3. Assessment of children and adolescents* (2nd ed.). New York: Wiley.

Finn, S. E. (1996, March). *Assessment feedback integrating MMPI-2 and Rorschach findings.* Paper presented at the annual meeting of the Society for Personality Assessment, Denver, CO.

Frank, L. K. (1939). Projective methods for the study of personality. *Journal of Psychology, 8*, 389–413.

Freedenfeld, R., Ornduff, S., & Kelsey, R. M. (1995). Object relations and physical abuse: A TAT analysis. *Journal of Personality Assessment, 64*, 552–568.

Freud, S. (1958). Psycho-analytic notes upon an autobiographical account of a case of paranoia (dementia paranoides). *Standard edition* (Vol. XII, pp. 9–82). London: Hogarth. (Original work published in 1911.)

Freud, S. (1962). Further remarks on the neuro-psychoses of defence. *Standard edition* (Vol. III, pp. 162–185). London: Hogarth. (Original work published in 1896.)

Ganellen, R. J. (1996). *Integrating the Rorschach and the MMPI-2 in personality assessment.* Mahwah, NJ: Erlbaum

Gardner, R. A. (1971). *Therapeutic communication with children. The mutual storytelling technique.* Northvale, NJ: Aronson.

Goodenough, F. L. (1926). *Measurement of intelligence by drawings.* New York: Harcourt, Brace & World.

Gresham, F. M. (1993). "What's wrong in this picture?": Response to Motta et al.'s review of human figure drawings. *School Psychology Quarterly, 8*, 182–186.

Haak, R. A. (1990). Using the sentence completion to assess emotional disturbance. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (Vol. 2, pp. 147–167). New York: Guilford Press.

Hammer, E. F. (1958). *The clinical application of projective drawings.* Springfield, IL: Charles C. Thomas.

Hammer, E. F. (1985). The House-Tree-Person test. In C. S. Newmark (Ed.), *Major psychological assessment instruments* (pp. 135–164). Boston: Allyn & Bacon.

Handler, L. (1985). The clinical use of the Draw-a-person test (DAP). In C. S. Newmark (Ed.), *Major psychological assessment instruments* (pp. 165–216). Boston: Allyn & Bacon.

Handler, L., & Habenicht, D. (1994). The Kinetic Family Drawing technique: A review of the literature. *Journal of Personality Assessment, 62*, 440–464.

Harris, D. B. (1963). *Children's drawings as a measure of intellectual maturity.* New York: Harcourt, Brace & World.

Hart, D. H. (1986). The sentence completion techniques. In H. M. Knoff (Ed.), *The assessment of child and adolescent personality* (pp. 245–272). New York: Guilford Press.

Hart, D. H., Kehle, T. J., & Davies, M. V. (1983). Effectiveness of sentence completion techniques: A review of the Hart Sentence Completion Test. *School Psychology Review, 12*, 428–434.

Haworth, M. (1965). *A schedule of adaptive mechanisms in CAT responses.* Larchmont, NY: CPS.

Hibbbard, S., Farmer, L., Wells, C., Difillipo, E., Barry, W., Korman, R., & Sloan, P. (1994). Validation of Cramer's defense mechanism manual for the TAT. *Journal of Personality Assessment, 63*, 197–210.

Hibbard, S., Hilsenroth, M. J., Hibbard, J. K., & Nash, M. R. (1995). A validity study of two projective object representation measures. *Psychological Assessment, 7*, 432–439.

Hoffman, S., & Kupperman, N. (1990). Indirect treatment of traumatic psychological experiences: The use of TAT cards. *American Journal of Psychotherapy, 44*, 107–115.

Holtzman, W. H. (1993). An unjustified, sweeping indictment by Motta et al. of human figure drawings for assessing psychological functioning. *School Psychology Quarterly, 8*, 189–190.

Hutton, J. B., Dubes, R., & Muir, S. (1992). Assessment practices of school psychologists: Ten years later. *School Psychology Review, 21*, 271–284.

Kahill, S. (1984). Human figure drawings in adults: An update of the empirical literature. *Canadian Psychology, 25*, 269–292.

Kennedy, M. L., Faust, D., Willis, W. G., & Piotrowski, C. (1994). Social-emotional assessment practices in school psychology. *Journal of Psychoeducational Assessment, 12*, 228–240.

Knoff, H. M. (1990). Evaluation of projective drawings. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (Vol. 2, pp. 89–146). New York: Guilford Press.

Knoff, H. M., & Prout, H. T. (1985a). The Kinetic Drawing System: A review and integration of the Kinetic Family and Kinetic School Drawing techniques. *Psychology in the Schools, 22*, 50–59.

Knoff, H. M., & Prout, H. T. (1985b). *The Kinetic drawing system: Family and school.* Los Angeles: Western Psychological Services.

Kohlberg, L. (1976). Moral stage and moralization: The cognitive developmental approach. In T. Lickona (Ed.), *Moral development and behavior: Theory, research, and social issues* (pp. 31–53). New York: Holt, Rinehart & Winston.

Koppitz, E. M. (1968). *Psychological evaluation of children's human figure drawings.* New York: Grune & Stratton.

Koppitz, E. M. (1984). *Psychological evaluation of human figure drawings by middle school pupils.* New York: Grune & Stratton.

Kuehnle, K. (1996). *Assessing allegations of child sexual abuse.* Sarasota, FL: Professional Resource Press.

Lah, M. I. (1989a). New validity, normative, and scoring data for the Rotter Incomplete Sentences Blank. *Journal of Personality Assessment, 53*, 607–620.

Lah, M. I. (1989b). Sentence completion tests. In C. S. Newmark (Ed.), *Major psychological assessment instruments, Volume II* (pp. 133–163). Boston: Allyn & Bacon.

Leichtman, M. (1996). *The Rorschach: A developmental perspective.* Hillsdale, NJ: Analytic Press.

Machover, K. (1948). *Personality projection in the drawing of the human figure.* Springfield, IL: Charles C. Thomas.

Machover, K. (1951). Drawing of the human figure: A method of personality investigation. In H. H. Anderson & C. L. Anderson (Eds.), *An introduction to projective techniques* (pp. 341–369). New York: Prentice-Hall.

McArthur, D. S., & Roberts, G. E. (1990). *Roberts Apperception Test for Children manual.* Los Angeles: Western Psychological Services.

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1953). *The achievement motive.* New York: Appleton-Century-Crofts.

McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review, 96*, 690–702.

McConaughy, S. H., & Achenbach, T. M. (1994). *Manual for the semistructured clinical interview for children and adolescents.* Burlington, VT: University of Vermont Department of Psychiatry.

McDowell, C., & Acklin, M. W. (1996). Standardizing procedures for calculating Rorschach inter-rater reliability: Conceptual and empirical foundations. *Journal of Personality Assessment, 66*, 308–320.

McNeish, T. J., & Naglieri, J. A. (1991). Identification of the seriously emotionally disturbed using the Draw A Person: Screening Procedure for Emotional Disturbance. *Journal of Special Education, 27*, 115–121.

Morgan, C. D., & Murray, H. A. (1935). A method for investigating fantasies: The Thematic Apperception Test. *Archives of Neurology and Psychiatry, 34*, 289–306.

Motta, R. W., Little, S. G., & Tobin, M. I. (1993). The use and abuse of human figure drawings. *School Psychology Quarterly, 8*, 162–169.

Murray, H. A. (1943). *Thematic Apperception Test manual.* Cambridge, MA: Harvard University Press.

Murray, H. A. (1951). Uses of the Thematic Apperception Test. *American Journal of Psychiatry, 107*, 577–581.

Murray, H. A. (1971). *Thematic Apperception Test: Manual* (Rev. ed.). Cambridge, MA: Harvard University Press.

Murstein, B. I. (1963). *Theory and research in projective techniques (Emphasizing the TAT).* New York: Wiley.

Murstein, B. I., & Wolf, S. R. (1970). Empirical test of the "levels" hypothesis with five projective techniques. *Journal of Abnormal Psychology, 75*, 38–44.

Naglieri, J. A. (1988). *Draw A Person: A quantitative scoring system.* New York: Psychological Corporation.

Naglieri, J. A., McNeish, T. J., & Bardos, A. N. (1991). *Draw-A-Person: Screening Procedure for Emotional Disturbance.* Austin, TX: ProEd.

Naglieri, J. A., & Pfeiffer, S. I. (1992) Performance of disruptive behavior disordered and normal samples on the Draw A Person: Screening Procedure for Emotional Disturbance. *Psychological Assessment, 4*, 156–159.

Obrzut, J. E., & Boliek, C. A. (1986). Thematic approaches to personality assessment with children and adolescents. In H. M. Knoff (Ed.), *The assessment of child and adolescent personality* (pp. 173–198). New York: Guilford Press.

Ornduff, S. R., Freedenfeld, R., Kelsey, R. M., & Critelli, J. (1994). Object relations of sexually abused female subjects: A TAT analysis. *Journal of Personality Assessment, 63*, 223–228.

Ornduff, S. R., & Kelsey, R. M. (1996). Object relations of sexually and physically abused female children: A TAT analysis. *Journal of Personality Assessment, 66*, 91–105.

Parker, K. C. H., Hanson, R. K., & Hunsley, J. (1988). MMPI, Rorschach and WAIS: A meta-analytic comparison of reliability, stability, and validity. *Psychological Bulletin, 103*, 367–373.

Piotrowski, C. (1996). The status of Exner's Comprehensive System in contemporary research. *Perceptual and Motor Skills, 82*, 1341–1342.

Piotrowski, C., & Keller, J. W. (1989). Psychological testing in outpatient facilities: A national study. *Professional Psychology, 20*, 423–425.

Prout, H. T., & Phillips, P. D. (1974). A clinical note: The kinetic school drawing. *Psychology in the Schools, 11*, 303–306.

Reynolds, C. R. (1978). A quick scoring guide to the interpretation of children's Kinetic Family Drawings (KFD). *Psychology in the School, 15*, 489–492.

Roback, H. B. (1968). Human figure drawings: Their utility in the clinical psychologist's armamentarium for personality assessment. *Psychological Bulletin, 70*, 1–19.

Ronan, G. F., Colavito, V. A., & Hammontree, S. R. (1993). Personal problem-solving system for scoring TAT responses: Preliminary validity and reliability data. *Journal of Personality Assessment, 61*, 28–40.

Rorschach, H. (1942). *Psychodiagnostics.* Bern, Switzerland: Hans Huber. (Original work published in 1921.)

Rotter, J. B., Lah, M. I., & Rafferty, J. E. (1992). *Manual—Rotter Incomplete Sentences Blank* (2nd ed.). Orlando, FL: Psychological Corporation.

Seaton, B., & Allen, J. (1996, March). *Interscorer reliability of Rorschach structural summary data.* Poster session presented at the annual meeting of the Society for Personality Assessment, Denver, CO.

Smith, D., & Dumont, F. (1995). A cautionary study: Unwarranted interpretations of the Draw-A-Person test. *Professional Psychology, 3,* 298–303.

Stinnett, T. A., Havey, J. M., & Oehler-Stinnett, J. (1994). Current test usage by practicing school psychologists: A national survey. *Journal of Psychoeducational Assessment, 12,* 331–350.

Stone, H. K., & Dellis, N. P. (1960). An exploratory investigation into the levels hypothesis. *Journal of Projective Techniques, 24,* 333–340.

Sutton, P. M., & Swenson, C. H. (1983). The reliability and concurrent validity of alternative methods for assessing ego development. *Journal of Personality Development, 47,* 468–475.

Swensen, C. H. (1957). Empirical evaluations of human figure drawings. *Psychological Bulletin, 54,* 431–466.

Swensen, C. H. (1968). Empirical evaluations of human figure drawings: 1957–1966. *Psychological Bulletin, 70,* 20–44.

Tharinger, D. J., & Stark, K. (1990). A qualitative versus quantitative approach to the Draw-a-Person and Kinetic Family Drawing: A study of mood- and anxiety-disorder children. *Psychological Assessment, 2,* 365–375.

Thomas, A. D., & Dudek, S. Z. (1985). Interpersonal affect in TAT responses: A scoring system. *Journal of Personality Assessment, 49,* 30–37.

Vane, J. R. (1981). The Thematic Apperception Test: A review. *School Psychology Review, 1,* 319–336.

Watkins, C. E., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology, 26,* 54–60.

Weiner, I. B. (1977). Projective tests in differential diagnosis. In B. B. Wolman (Ed.), *International encyclopedia of neurology, psychiatry, psychoanalysis, and psychology* (pp. 112–116). Princeton, NJ: Van Nostrand Reinhold.

Weiner, I. B. (1986). Assessing children and adolescents with the Rorschach. In H. M. Knoff (Ed.), *The assessment of child and adolescent personality* (pp. 141–171). New York: Guilford Press.

Weiner, I. B. (1993). Clinical considerations in the conjoint use of the Rorschach and the MMPI. *Journal of Personality Assessment, 60,* 148–152.

Weiner, I. B. (1994). Rorschach assessment. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome evaluation* (pp. 249–278). Hillsdale, NJ: Erlbaum.

Weiner, I. B. (1995a). Methodological considerations in Rorschach research. *Psychological Assessment, 7,* 330–337.

Weiner, I. B. (1995b). Psychometric issues in forensic applications of the MMPI-2. In Y. S. Ben-Porath, J. R. Graham, G. C. N. Hall, R. D. Hirschman, & M. S. Zaragoza (Eds.), *Forensic applications of the MMPI-2* (pp. 48–81). Thousand Oaks, CA: Sage.

Weiner, I. B. (1996). Some observations on the validity of the Rorschach Inkblot Method. *Psychological Assessment, 8,* 206–213.

Weiner, I. B. (1997). Current status of the Rorschach Inkblot Method. *Journal of Personality Assessment, 68,* 5–19.

Weiner, I. B., & Exner, J. E., Jr. (1991). Rorschach changes in long-term and short-term psychotherapy. *Journal of Personality Assessment, 56,* 453–465.

Westen, D., Klepser, J., Ruffins, S. A., Silverman, M., Lifton, N., & Boekamp, J. (1991). Object relations in childhood and adolescence: The development of working representations. *Journal of Consulting and Clinical Psychology, 59,* 400–409.

Westen, D., Lohr, N., Silk, K., Kerber, K., & Goodrich, S. (1985). *Object relations and social cognition TAT scoring manual.* Ann Arbor, MI: University of Michigan.

Westen, D., Ludolph, P., Block, M. J., Wixom, J., & Wiss, F. C. (1990). Developmental history and object relations in psychiatrically disturbed adolescent girls. *American Journal of Psychiatry, 147,* 1061–1068.

Westen, D., Ludolph, P., Lerner, H., Ruffins, S., & Wiss, F. C. (1990). Object relations in borderline adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 29,* 338–348.

Worchel, F. F., Rae, W. A., Olson, T. K., & Crowley, S. L. (1992). Selective responsiveness of chronically ill children to assessments of depression. *Journal of Personality Assessment, 59,* 605–615.

F.P.C.E. BIBLIOTECA U.P.

4.16 Assessment of Schema and Problem-solving Strategies with Projective Techniques

HEDWIG TEGLASI
University of Maryland, College Park, MD, USA

4.16.1 INTRODUCTION 460

4.16.2 PATTERNS OF USE 460

4.16.3 BASIC ASSUMPTIONS OF PROJECTIVE TECHNIQUES AND RESEARCH FROM OTHER PSYCHOLOGY SUBFIELDS 461

4.16.3.1 Schema Theory and Projective Techniques 462
4.16.3.2 Unconscious Processing of Information 464
4.16.3.3 Inert Knowledge vs. Usable Knowledge 465
4.16.3.4 Affective and Motivational Influences on Cognition 465
4.16.3.5 Narrative Psychology and Thematic Apperception Techniques 466

4.16.4 LEVELS OF PERSONALITY ASSESSED BY PROJECTIVE AND QUESTIONNAIRE METHODS 467

4.16.5 SELF-REPORT VS. PROJECTIVE PERSONALITY TESTING 470

4.16.6 CONTRIBUTION OF PROJECTIVE TECHNIQUES TO DIAGNOSIS AND INTERVENTION 470

4.16.7 PROJECTIVE TECHNIQUES AS PERFORMANCE MEASURES OF PERSONALITY 472

4.16.7.1 Total Personality 474
4.16.7.2 Many Correct Solutions 476
4.16.7.3 Generalize to Different Criteria 476
4.16.7.4 Differences in Conditions of Learning and Performance 477

4.16.8 SPECIFIC PROJECTIVE TECHNIQUES 477

4.16.8.1 Stimulus 477
4.16.8.2 Response 477
4.16.8.3 Task Demand 478
4.16.8.4 Interpretation 478

4.16.9 THEMATIC APPERCEPTIVE TECHNIQUES 479

4.16.9.1 Response 479
4.16.9.2 Stimuli 479
4.16.9.3 Interpretation 480

4.16.10 RORSCHACH TECHNIQUE 481

4.16.10.1 Stimulus Features and Response Parameters 481
4.16.10.2 Interpretation 483

4.16.11 DRAWING TECHNIQUES 482

4.16.11.1 Interpretation 484

4.16.12 CASE ILLUSTRATION 484

volume

1

Basic Foundations and Principles of Interpretation

4th edition

THE RORSCHACH®

A Comprehensive System



JOHN E. EXNER JR., PHD

CHAPTER 11

The Response Process

Anyone using the Rorschach should be aware of how responses are formulated, and how those that are delivered create a basis from which information about the psychology of a person can be gleaned. This information is important because it provides the framework for understanding interpretive principles and procedures. To the uninformed, it may seem incredulous that the responses to 10 inkblot-like figures can provide much information about an individual. It is an issue that has often provided grist for the mill of the skeptic and, unfortunately, some explanations about how the test works have only increased that skepticism.

This skepticism probably increased most when the projective movement reached its zenith and the Rorschach was identified as the centerpiece among "projective techniques." During that time, major emphasis was on content analysis, and the nature of the test was usually conceptualized in terms of the projective process. Some attempted a direct symbolic interpretation of specific kinds of contents (Phillips & Smith, 1953). Others sought to equate the blot stimuli with universal symbolic meanings, generating faulty notions about a father card, mother card, sex card, interpersonal card, and so on (Halpern, 1953; Meer & Singer, 1950; Pascal, Ruesch, Devine, & Suttell, 1950). Although emphasis on the projective process served to enhance the usefulness of the test, it also detracted from a broader understanding of the nature of the test, and tended to detract from research on Rorschach's original thinking about the method.

RORSCHACH'S CONCEPT OF THE METHOD

The process of the test intrigued Rorschach, and he did formulate some hypotheses related to the response operations. He postulated that responses are formed through an integration of memory traces with the visual images created by the stimulus figure. He argued that the effort to match the stimulus images with existing engrams is *consciously realized.* In other words, the person is aware that the blot is not identical to objects stored in memory. Consequently, the method requires a willingness by the individual to identify the blot, or blot area, as being something that it is not, but to which it has some similarity.

Rorschach described this as an associational process and postulated that differences in "thresholds" exist among people for their ability to assimilate or integrate the stimulus sensation with the existing engrams. He believed that the differences in threshold were the main cause for the broad array of responses that occurs. It was on this premise that he rejected the notion that unconscious elements might be influential in forming a response. He viewed the response process as one of perception and/or apperception.

Rorschach also argued, quite persuasively, that imagination had little or nothing to do with the *basic* process of the test, but that it could manifest in embellishments to responses. It seems likely that if Rorschach had lived long enough to consider Murray's concept of projection (1938),

or Frank's (1939) formulation of the *Projective Hypothesis,* he would have acknowledged them as relevant to decisions leading to the responses. But, it is equally likely that he would have disavowed the concept of projection as representing a major component in the response process.

There is no question that the nature of the test, and of the response process, were sorely neglected areas of research for several decades following Rorschach's death. This neglect probably contributed substantially to the broad divergence that occurred among those attempting to develop the test. Interestingly, the accumulation of research findings lends support for most of Rorschach's postulates about the response process, although it appears to be considerably more complex than he conceived. Numerous operations occur before a response is actually delivered, and they occur within a brief time frame that few recognized during the early days of the development of the test.

THE RORSCHACH AS A DECISION-MAKING TASK

Rorschach was correct in his assumption that the response is formulated with an awareness that the blot is not identical to existing memory traces. Usually, the person has some awareness of this well before the first blot is exposed, by reason of the introduction of the task by the examiner. Cattell (1951), in his effort to describe the projective situation, probably best described the task of the person who is taking the Rorschach for the first time. He noted that the individual is required to identify something that is not actually there. According to Cattell, the task requires the person to "misperceive" the stimulus and through that misperception the person is encouraged to project something of himself or herself into the response.

Although Cattell may be correct in describing the projective situation, that description can be quite misleading when applied to the circumstances of the Rorschach, depending on whether the term *misperceive* is translated to mean how the stimulus is *translated,* or how the stimulus is *identified.* This is a very important distinction, because the first implies that the person neglects the fact that the stimulus is an inkblot, whereas the second does not carry that implication. Rorschach believed that identification is one of the essential operations in the response. When he occasionally encountered persons who would concretely name the blots, he concluded that they were so seriously debilitated by reason of intellectual deficits, neurological impairment, or active psychosis, that their associational or integrative operations failed or decayed. Most professionals experienced with the Rorschach will have encountered individuals who are so impaired that they simply cannot respond to the task, or who are so detached from reality that they offer only hallucinatory type of responses when confronted with a blot (e.g., "My God, that sounds terrible, take it away."). But these are very rare instances. The overwhelming majority of those who take the test are fully aware that they are responding to an inkblot.

Examiners alert the adult by saying something such as, "Now we are going to do the inkblot test" or, as is common when testing young children, "Now we're going to look at some inkblots." But even if the introduction fails to include the use of the word *inkblot,* anyone would be hard-pressed to argue that the person is not aware of the fact. For instance, when the records of 500 patient and nonpatient 5-, 6-, and 7-year-olds were reviewed, it was found that 207 gave comments or responses at the onset implying an awareness of the nature of the stimulus (Exner, 1980). These varied from, "It's an inkblot," or "It's a bunch of ink," to comments such as, "I know how to make these," or "We make prettier ones than this." In reality, those children gave the only truly correct answer. The stimulus is only an inkblot! But if that correct answer is delivered as the *first* response, it is not accepted. Instead, the examiner encourages some other identification, usually by saying, "Yes, I know. This is the inkblot test, but what might it be?"

In effect, the nature of the test situation forces the person to identify the figure as something it is

not. In some ways, the Rorschach is a *problem-solving* situation because there is a requirement to violate reality while maintaining one's personal integrity. The requirement to *misidentify* the stimulus provokes a complex set of psychological operations that ultimately culminates in decisions that lead to the delivery of answers.

DECISION CHOICES

The problem posed by the need to misidentify the stimulus would be simple for most people if only one alternative existed to the answer *inkblot,* but that is not the case. Many potential answers, consistent with the figure or parts of it, are available. Most people recognize this quickly after the figure is presented. Thus, one decision that the person must address while taking the test is which of the possible answers to verbalize, and which to discard.

For several decades after the test was published, most who used and researched it remained unaware of the substantial frequency of potentially reasonable answers that seem to be readily available to most people. There are probably several reasons for this, beginning with Rorschach's rather sketchy report of his experiment, and especially how the figures were created. His failure to emphasize the exquisite detailing of the figures created for the test, tended to reinforce the assumption that the figures are largely ambiguous, an assumption that has proven to be false (Exner, 1996). Nonetheless, some of the information that Rorschach reported about his investigation seemed to support the notion of ambiguity, and led those working to develop the test into a direction that, generally, ignored the stimulus properties of the figures.

Rorschach's observations led most of those involved in developing the test to form numerous interpretive hypotheses about short and long records, and about reaction time data.[1] The latter became particularly widespread across the several systems. All of the systematizers, and many other researchers of the test, were quick to incorporate and expand Rorschach's concept of color shock, and also to formulate the notion of "gray-black shock." Both were postulated to equate with forms of anxiety, and were thought to represent neurotic characteristics (Beck, 1945; Klopfer & Kelley, 1942; Miale & Harrower-Erikson, 1940; Piotrowski, 1957). The implication was that the person was somehow traumatized by the blot features, and therefore struggled to form a response.

Another factor that tended to mislead most Rorschachers about the availability of potential answers has been the variety of published norms. Although they have varied depending on the system employed, the mean number of responses for adults has generally ranged from about 22 to 32 responses, with standard deviations ranging from five to eight (Beck, Beck, Levitt, & Molish, 1961; Exner, 1974, 1978). The data concerning children yielded even lower means for *R* (Ames, Learned, Metraux, & Walker, 1952; Ames, Metraux, & Walker, 1971; Beck et al., 1961; Exner, 1978; Exner & Weiner, 1982, 1995). These data suggest that the average person might find, or misidentify, two or three objects per blot, and few would have argued with the position that some people have difficulty formulating more than one answer to some of the blots. That conclusion has been afforded further support by hundreds of research studies in which the average number of responses given generally falls within the "normal" range.

---

[1] Rorschach noted that most of his subjects gave between 15 and 30 responses, with depressed, "sullen or unobliging" subjects tending to give the fewest. He also noted that most people delay before giving answers, and suggested that those who gave several responses very quickly were probably "scattered" in their perception or ideation. He did not record reaction times for first responses, or total times per blot, but his notion of "scattered," plus another hypothesis concerning *color shock,* caused those who followed him to stress the faithful recording of those times. He used the term *color shock* to describe instances in which a person appeared to have considerable difficulty in forming a response to Card VIII, the first totally chromatic blot, although responding to the prior blots at a seemingly natural rate. He speculated that this apparent helplessness indicated some form of emotional suppression.

Another element contributing to the false notion that people formulate only a small number of answers per blot is the occasional instances in which people report that they cannot find or see anything other than the blot itself. Rorschach noted that some of his subjects did give "refusals" and suggested that this might result from an insurmountable blocking process. Klopfer and Kelley (1942) tended to agree with that postulate, but Beck (1945) argued that the rejection, or tendency to reject, might also be caused by problems in the perceptual organizing process, especially for the more difficult blots. This issue stimulated several studies on card difficulty, in which *reaction time* was used as an index of the difficulty or complexity of the stimuli (Dubrovner, VonLackum, & Jost, 1950; Matarazzo & Mensh, 1952; Meer, 1955; Rabin & Sanderson, 1947). Meer used a transformation of reaction times and form accuracy data from 12 studies to identify blot difficulty levels. Frequency data collected at the Rorschach Research Foundation suggest that most of his conclusions were correct concerning blot difficulty and/or complexity levels (Exner, Martin, & Cohen, 1983). However, findings such as those reported by Meer have only served to reinforce the flawed notion that people often have to struggle to find more than one object in each blot. Few practitioners, or researchers, have entertained the notion that most people form potential answers quickly and with relative ease, yet this fact appears to be supported by a series of studies.

THE RANGE OF POTENTIAL RESPONSES

An issue of concern during the development of the Comprehensive System was the impact of the examiner on the subject. A considerable literature had evolved, suggesting that some features of Rorschach responses can be altered under conditions that vary from the standard procedures. For example, different instructional sets, such as asking the person to find more things, find things moving, find more small objects, and so on, will usually produce more of the kinds of responses for which the set is established (Coffin, 1941; Hutt, Gibby, Milton, & Pottharst, 1950; Abramson, 1951; Gibby, 1951).

Similarly, differences among the instructions used by the various Rorschach Systematizers produce substantial differences in the average length of a record (Klopfer = 23.9; Beck = 31.2; Hertz = 32.9; Piotrowski = 33.8; Rapaport = 36.4). Goetcheus (1967) used 16 examiners in a crossover design to study the differences between the Beck and Klopfer instructions. Each examiner administered eight tests using the Beck instructions, and eight using the Klopfer instructions. She found that the Beck instructions, which include, *Tell me everything you see,* and prompt for a second answer on the first five cards if only one response is given, produced records that, on average were six responses longer than those administered by the Klopfer format.

Several studies have also demonstrated that reinforcement, both verbal and nonverbal, can alter the frequencies of some kinds of responses (Wickes, 1956; Gross, 1959; Dinoff, 1960; Magnussen, 1960; Hersen & Greaves, 1971). As it turns out, all or most of these designs were tapping into the fact that most people generate many potential responses when confronted with the inkblot figure. Although some hints of this fact can be culled from some of the early research on the response process, no one did so. It was not until the 1970s that an awareness of this occurred somewhat serendipitously.

In a pilot investigation concerning the effects of reinforcement, two groups of 10 persons each, one of nonpatients and the second of outpatients, were instructed to give as many responses as they could, with a time limit of 60 seconds per blot (Exner & Armbruster, 1974). They were reinforced with a payment of 10 cents per response, paid immediately as each response was delivered. The nonpatient group averaged 104 answers to the 10 figures, with a range of 68 to 147 responses. The outpatients averaged 113 responses, with a range of 71 to 164 answers. These unexpectedly

large numbers raised several questions. First, to what extent did the reinforcement of the dimes alter the Rorschach response process? Second, to what extent did the exposure time of 60 seconds force people to rescan the field, thereby forming responses that might not have been formulated under the standard administration conditions. Third, did the reinforcement condition cause people to violate the use of accurate form more frequently than might be typical? Finally, because only one examiner was used in the pilot, would people give unusually large numbers of responses if several examiners were used? These questions led to the design of a more sophisticated study.

Exner, Armbruster, and Mittman (1978) used 12 experienced examiners to administer the test to five groups of 20 subjects each. Groups 1 and 2 consisted of 40 adult nonpatients, ranging in age from 20 to 41 years, subdivided into groups of 20 each by a median split of their distribution of scores on the K Scale of the MMPI. Group 3 consisted of 20 nonpatient children, ranging in age from 11 to 13 years. Group 4 was comprised of 20 inpatient depressives, aged 29 to 51 years, and Group 5 consisted of 20 inpatient schizophrenics, ranging in age from 24 to 42 years. None of the participants had been administered the Rorschach previously. The examiners were randomly assigned so that none tested more than four persons from a single group or more than 10 persons in all. All of the subjects had volunteered to participate in a *standardization* study concerning the inkblot test. The procedure followed the standardized method except that, prior to the beginning of the test, each person was told that he or she would have each blot for 60 seconds, and during that time, they should report as many things as they could find in the figure.

The responses were audio recorded, and a silent signal was entered on the tape at 15-second intervals. The examiners would rewind the tape after the responses to all 10 cards had been given, and play them back, one at a time, to inquire only for location. This permitted a review of the answers for the appropriate use of form. The average number of responses given by each group is shown in Table 11.1, which also shows the average number given during the first 15-second interval, the second 15-second interval, and the last 30 seconds of exposure, plus data concerning the mean *X+%* and the mean frequency of Popular responses for each interval.

The data in Table 11.1 indicate that, in the experimental condition, all five groups gave two to four times the average of 22 answers usually obtained using the standard administration procedure. The data are even more intriguing when the average number of responses given during the first 15 seconds of blot exposure are examined. The three nonpatient groups all gave at least one-third more answers during that interval than is usually the case under the standard administration conditions, during which most people retain the cards for between 30 and 50 seconds. Both psychiatric groups gave at least as many answers during the first 15 seconds of exposure as they would be expected to give when the test is administered under standard conditions.

In addition to the finding that people can give many answers when instructed to do so, the data regarding the appropriate use of form are important. The *X+%* represents the proportion of common or conventional answers in a record, whereas Populars are the most common responses.[2] The means for the *X+%* indicate that each of the four nonschizophrenic groups were relatively consistent in giving common responses throughout the 60-second response period. They did not violate form accuracy markedly during any of the intervals of exposure to the blots. The data for Popular answers are also important in this context. All four of the nonschizophrenic groups averaged about eight Popular answers during the first 30 seconds of exposure, and about two-thirds of those occurred during the first 15 seconds of exposure to the figures. Thus, in a relatively brief exposure

[2] Although 13 responses currently are identified as Popular in the Comprehensive System, 17 answers were being coded as *P* at the time of the study.

Table 11.1 Average Number of Responses and Means for X+% and Popular Responses for Each of Five Groups during Four Intervals of Blot Exposure.

	First 15 Seconds		Second 15 Seconds		Second 30 Seconds		Total 60 Seconds	
	M	SD	M	SD	M	SD	M	SD
20 Nonpatients, upper half MMPI K scale								
R	30.4	4.1	31.2	5.8	21.7	4.1	83.3[a]	9.2
X+%	88.9	11.1	81.4	9.9	89.3	9.6	85.1	10.2
P	5.2	2.1	3.5	1.3	3.2	1.1	10.8	2.8
20 Nonpatients, lower half MMPI K scale								
R	38.1	6.8	32.2	6.1	30.4	7.8	100.6	10.4
X+%	83.2	9.7	79.6	7.6	78.1	8.7	79.9	9.8
P	5.0	1.9	2.4	1.1	1.9	0.9	9.3	3.1
20 Nonpatient children								
R	38.9	7.1	30.7	4.3	24.5	8.3	94.1	9.8
X+%	84.6	7.8	80.1	8.5	84.1	7.8	83.3	8.1
P	5.3	1.8	2.4	1.1	2.0	1.1	9.7	2.2
20 Inpatient schizophrenics								
R	22.7[b]	6.2	18.1[b]	5.1	22.4	6.7	63.2[b]	9.4
X+%	63.2[c]	10.8	54.6[c]	11.7	49.3[c]	11.4	53.6	12.7
P	2.4[c]	1.6	1.7	1.0	4.3[c]	1.7	8.4	3.8
20 Inpatient depressives								
R	14.8[b]	4.4	17.1[b]	5.7	19.3	7.8	51.2[b]	7.8
X+%	77.1	6.8	72.3	7.1	68.7	8.3	71.9	8.9
P	6.2	3.1	3.1	1.4	0.9	0.7	10.2	4.3

[a] Significantly less than nonpatients in the lower half of the MMPI K Scale distribution, $p < .01$.
[b] Significantly different from nonpatient groups, $p < .01$.
[c] Significantly different from all other groups, $p < .01$.

time, the people in those four groups gave as many, or more, *P* answers than is customary when the test is given under standard conditions.

Most people, especially the nonpatients, began responding very quickly to each figure, usually within two or three seconds. There is a considerable variance for the average number of answers per card. The most solid figures, IV, V, VI, and IX, generally yielded the lowest average number of answers, whereas the more broken figures, III, VIII, and X, yielded the highest average number of answers.

The results of this study indicate that people can form multiple potential responses, generally congruent with the contours of the stimulus, in a relatively brief interval after the figure is exposed. The shortest nonpatient record contained 56 answers, and the shortest patient record (from a depressed person) contained 34 responses. When only the number of answers given during the first 30 seconds is considered, the lowest number of answers from a nonpatient is 23, and the highest is 89. One depressed patient gave only 10 responses during the first 30 seconds, but 32 of the 40 psychiatric subjects gave at least 16 responses in that interval.

These findings differ markedly with most reference data that have been published for both patients and nonpatients. For instance, in the United States, the mean *R* for nonpatient adults is 22.3 responses, with the majority of records falling between 17 and 27 answers. The average *R* for samples collected in other countries is reasonably similar. For instance, data for nonpatient samples from several countries were reported in a symposium at the XVI Congress of the International

Rorschach Society (Erdberg & Shaffer, 1999). The samples involved included *N*'s ranging from 72 to 520 subjects. Some of the mean *R*'s reported are: Belgium = 22.7; Finland = 22.1; Denmark = 24.5; Japan = 26.3; Peru = 22.0; Portugal = 22.1; Spain = 24.8.

THE RESPONSE PROCESS

If people can easily form many potential answers to each figure, why is it they deliver far fewer than are available when the test is administered in the standardized manner?

Extrapolating from the Exner, Armbruster, and Mittman study, it seems reasonable to suggest that most people will give between 25% and 35% of the potential answers that they formulate. Any attempt to understand why this occurs must be framed in the context of the psychological activities that occur when forming and delivering responses.

In the very few seconds before a response is given, at least six mental operations occur. They are not discrete, but there may be some merit in considering them as three distinct phases of the response process.

Phase I:
1. Visual input and encoding of the stimulus and its parts.
2. Classification (identification) of the stimulus and/or its parts and a rank ordering of the potential responses that are created.

Phase II:
3. Discarding potential answers that have low rankings.
4. Discarding other potential answers through censorship.

Phase III:
5. Selection of some of the remaining responses by reason of traits or styles.
6. Selection of some of the remaining responses because of state influences.

It is after these operations have occurred, that the first answer to a figure is given. The process is activated by the examiner's query, "What might this be?" In some ways, the process is similar to that which takes place if a person is shown a photo of a chair and asked the same question. Although the instructions are brief, there is an implication that the task is to select a response that is commensurate with the *distal properties* of the stimulus field. Distal properties are the true components of a stimulus field. For instance, a chair has legs, a seat, and a back. It may have other features, but those elements, sometimes referred to as *critical distal bits* (Attneave, 1954), differentiate the chair from other objects that have similar distal features, such as a table or stool.

Critical distal bits are the most potent distal properties of a field. In effect, they create the parameters that limit the range of possible identifications of an object and constitute the elements that encourage specific identifications. For example, most people easily differentiate a baseball from an orange even though both are spherical and approximately the same size. The differentiation is made easy because both have critical distal properties that are constant to the object. The orange has a rough texture and a distinct coloring, whereas the baseball is white and has seams.

The task of reconciling the features of a Rorschach figure with true objects is made more complicated because the distal properties of the figure are not as precise and discrete as those of a chair or baseball. However, as noted earlier, Rorschach created the figures of the test in a way that ensured each contain numerous distinctive features that could easily be identified as being similar to objects stored in the memory traces of the individual. Thus, the figures fall far short of being completely or largely ambiguous. Each contains critical distal bits that create parameters limiting the array of translations that will be congruent with the distal properties of the field or parts of it (Exner, 1996). The presence of these discrete features facilitates the relatively rapid formation of potential answers.

THE INPUT PROCESS

It is difficult to appreciate the capacity of the human being to process information. The Rorschach figures are a form of visual stimulation

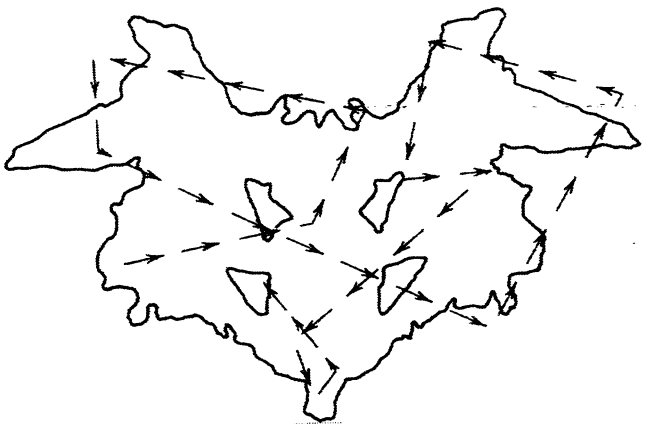


Figure 11.1 Visual scanning by a 19-year-old female during the first 500ms of blot exposure.

and the processing of visual information occurs quite rapidly. Although theories of visual processing remain open to discussion (Hochberg, 1981; Neisser, 1976; Pomerantz & Kubovy, 1981), an extremely large number of studies have demonstrated that pattern and/or picture recognition can occur very quickly (Fisher, Monty, & Senders, 1981). Although the number of studies concerning visual processing has increased dramatically during the past three decades (probably as a result of increasingly sophisticated technology), the methodology was not applied to Rorschach research until recently. Exner (1980, 1983) studied the visual scanning activity of nonpatient adults to some of the Rorschach blots. Figure 11.1 is a crude facsimilie of the eye-scanning activity of a 19-year-old female viewing Card I for approximately 500 ms.[3]

The arrows represent the focal point of the visual field of the individual. It is not certain how far the peripheral visual field extends, but if a one inch area on either side of the centerpoint is used as a conservative estimate, it would appear that this woman has viewed all of the blot, and some parts more than once in the course of one-half second. It seems clear that most everyone would scan Card I very completely in the course of one full second. The importance of this finding is that, for 125 adult nonpatients, the average reaction time for the first answer to Card I, measured by a voice actuated timer, is 5.79 seconds ($SD = 2.38$). In other words, even if one second is required to scan and encode the entire figure, nearly five seconds remain before the first answer is usually given.

Figure 11.2 is a crude facsimilie of the scanning action of a 23-year-old nonpatient male when shown Card III. It reflects his eye activity during the interval of 1.1 seconds after the figure was presented. He scanned this card at about the same pace as other persons in the study. His eye activity indicates that he has viewed all of the features of this broken blot at least once, and some segments more than once during this interval. The average reaction time for the first response to this card is 7.74 seconds ($SD = 3.1$), indicating that the majority of people offer their first answer in a period of about 4.5 to nearly 11 seconds. Thus, it seems reasonable to suggest that an interval of five to nine seconds elapses *after* the input occurs until the first answer is given.

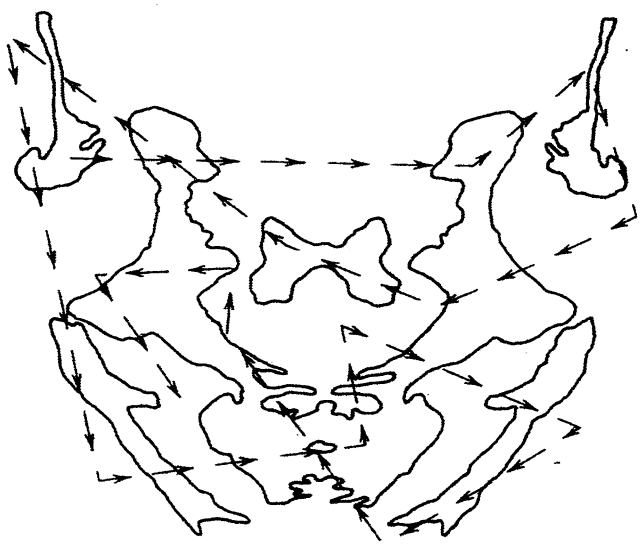


Figure 11.2 Visual scanning by a 23-year-old male during the first 1100ms after blot exposure.

[3] A Gulf & Western Model 200 Eye Movement Monitor was used to record the scanning activity. The subject's head was held in a retainer to minimize random head movement. The blots were presented tachistoscopically on a small screen in the center of the subject's visual field. The eye movements were recorded through infrared sensors attached to spectacle frames worn by the subject and transmitted in an analog model to a computer and a digital model to a converter, which reproduced the activity on a video display.

The importance of these findings is not that the person is able to input the stimulus quickly. That result is expected based on many studies about eye activity. The importance rests in confirmation of the *delay period* that exists between the input and the output. It is in that critical interval of a few seconds that many substantive operations take place that lead to the decisions of how the person will use the potential answers that he or she has identified after being exposed to the stimulus.

THE CLASSIFICATION (IDENTIFICATION) PROCESS

As the stimulus field is scanned, it is encoded and held in a form of short-term storage, and the process of classification begins. Data from long-term storage are used as the basis for comparison to classify (identify) the field and/or its parts. It is also important to note that the encoding and storage of the image does not cease after a brief interval. The person taking the test retains the card throughout the response period. Visual scanning continues during this time and it is likely that the encoding becomes more precise.

As classification begins, the presence of distinctive or critical distal features in the encoded field becomes quite important. In some instances, a person may not easily classify the field as a whole, or some of its parts, because of their apparent ambiguity. But almost all people, with the possible exception of those with severe intellectual or neurological deficits, will easily identify some elements in the field as being sufficiently congruent with known or imagined objects to create potential answers. This is because the critical distal features, in each of the 10 fields, create a similarity to real or imagined objects. As a consequence, some whole blots, or blots areas, have a greater likelihood of being included among the responses that are actually delivered when the rank ordering and discarding procedures are completed.

This occurs because of the way that Rorschach designed each figure. He did not simply drop paint or ink on a page and fold it to create symmetry and settle for the result. That is how he began but, after creating an ambiguous figure, he carefully sketched the details of the blot. His objective was to ensure that each blot contained distinct features that would tend to promote specific classes of responses. The presence of Popular responses probably best illustrates the intent of his effort.

Popular Responses

Although Popular answers are defined as those responses that occur in at least one out of every three protocols, that criterion is very conservative. In reality, most Popular responses appear much more frequently. For instance, more than 90% of all subjects identify the *D1* area of Card VIII as some sort of four-legged animal. This is because the contours of that area are very comparable to the legs, body, and head of a four-legged animal, and its location is somewhat discrete in the figure. Thus, even though the coloring of this area (pink) is incongruous to a four-legged animal, its distinctiveness and location, plus the fact that other areas of the blot are less easily identified, promote a high frequency of animal responses.

Similarly, approximately 85% of all subjects identify Cards I and V as a winged object, usually a bat or butterfly, and more than 80% identify the *D9* area of Card III as a human figure. In fact, only two of the 13 Popular answers appear in fewer than 50% of the 7,500 protocols that were used as the data set from which to distinguish Populars. The human or human-like figures response to the *D3* area of Card IX appear in only 39% of the records and the animal heads or whole animals to the *D1* area of Card II occur in only 35% of the protocols.

Competing Critical Bits

In addition to the primary distal features that promote Popular answers, there are other distal properties in each of the 10 figures that seem quite potent in contributing to frequent classifications of objects or classes of objects. Sometimes,

more than one identification of a blot field is encouraged because of competing critical distal features. Findings regarding Cards I and V are good examples.

Exner (1959) changed the grey-black coloring of Card I to light blue, yellow, or pink, holding all other features constant. He found that when the figure is blue or yellow, the Popular bat response, usually appearing in about 45% of protocols, declines to an almost zero frequency. Conversely, the Popular butterfly answer, usually appearing in about 40% of records, was given to more than 80% of the chromatically colored versions of Card I. Thus, it is realistic to surmise that the achromatic color is important to the identification of the field as a bat. It has also been found that if the *Dd34* projections are eliminated from the Card I figure, a marked decline in the frequency for both the bat and butterfly answers occurs, and an increase in face or mask responses occurs. A similar manipulation of response frequencies for Card V responses results when the distal features are altered.

About 85% of persons taking the test identify Card V as a bat or a butterfly. The proportions for each answer are nearly the same. About 44% give the butterfly answer and about 41% give the bat response. Why do some people identify the blot as a bat while others translate it as a butterfly? The difference is not attributable to gender, age, types or severity of pathology, or basic response styles. The differentiation apparently occurs as a result of the weight that individuals give to one or more of three critical distal bits in the Card V stimulus field (Exner, 1996). One is the grey-black coloring of the figure, which is somewhat incongruous with the butterfly answer. When the color is altered to pink, a change in response frequencies does occur, although not as much as might be expected. About 55% of respondents give the butterfly answer, but about 30% continue to give the bat response (Exner, 1959).

The other two critical bits in the Card V field are contours, and they tend to have more potency for those selecting either the bat or butterfly responses. The first is the side *D10* projections, which seem to be incongruous with both answers, but apparently are not. When the *D10* area is eliminated from the blot approximately 70% of respondents give the butterfly answer whereas only about 15% give a bat response. Obviously, the *D10* area is critical, either serving to thwart the butterfly answer or facilitate the bat answer. Another critical area is the top *Dd34* projections. When *Dd34* is deleted, about 70% of respondents identify this modified field as a bat and fewer than 15% report a butterfly. Clearly, the *Dd34* area is an important distal feature to those who select the butterfly answer (Exner, 1996).

Distal Feature Potency

The potency of distal features will often change if their position in the visual field is altered. For example, when Card VII is viewed in the upright position, about 65% of respondents identify the *D2* area as a human figure and about 25% identify it as an animal, usually a rabbit. When Card VII is viewed sideways, the *D2* area is rarely identified as a human figure and about 50% of respondents identify it as a dog (Exner, 1996).

Although contours probably represent the most important critical distal features in the blots, the color and shading properties sometimes play a very potent role in establishing response parameters. For instance, if Card VI is presented as solid grey, so that the internal shading variations are eliminated, the frequency of the Popular animal skin or pelt response to the *D1* area is reduced very significantly. However, if the shading is held constant and the color is chromatic (blue, yellow, or pink), the frequency of the Popular response does not decline (Exner, 1961).

Sometimes, a distal feature of a blot is influential to the formation of an answer even though it may not be specifically articulated in the response. For example, when the Popular animal response is given to the *D1* areas of Card II, movement (*FM*) is included as a determinant in about 85% of those answers. The animals are identified in playful

activity (COP) in about 30% of those responses, and in relation to aggressive activity, usually fighting (AG) or hurt (MOR), in nearly 40% of the answers. When a COP is given, the red areas of the blot (usually *D3*) are articulated in about two of every five answers. When an AG or MOR response is given, the red areas (either *D2* or *D3*) are articulated in about three of every five responses.

If the red *D2* and *D3* areas are deleted from the blot, *or* if they are changed in color to grey-black, the frequency of animal movement responses remains about the same but the proportion of AG and MOR movement answers is reduced to less than 5%, while the proportion of COP answers increases to nearly 70%. It seems obvious that the presence of the red areas in the Card II field contributes substantially to the formation of the AG and MOR answers even though the red *is not* articulated in about 40% of those responses.

Sometimes distal features serve to call attention to an area, and by doing so, increase the response rate to that area. For instance, the most common answers given to the blue-colored *D1* area in Card X are spider or crab. Exner and Wylie (1976), pursuing the hypothesis that the blue creates a dissonant effect, used a dye-coupling technique to alter the color of the *D1* area to a reddish-brown. The entire test was administered to two groups of 50 subjects each, one of nonpatients and one of inpatients, with a random half in each group being administered the standard Card X and the other half the altered version. The groups administered the altered version of Card X gave significantly fewer spider and crab responses than did the controls, even though the altered coloring was more congruent with either of those two answers.

This finding seemed puzzling until it was noted that the persons in the experimental group gave almost twice as many answers than controls to another area of the blot, *D6,* which is colored blue. A 30-item questionnaire was hastily contrived and administered to 50 people who had participated in the experiment. It contained one critical item, *What is your favorite color,* to which 41 persons responded with the answer, "blue." Apparently, the blue coloring of the area causes more people to attend to it during the classification process. Exner and Wylie (1977) created a completely achromatic version of Card X to test this hypothesis. They administered the entire test to 30 nonpatient volunteers, half of whom were administered the standard version of Card X and half the achromatic version. Only three subjects in the experimental group gave spider or crab responses to the *D1* area, whereas 12 of the 15 persons in the control group did so. Similar findings for Card X have been reported by Silva (2002), who has compared responses to chromatic and achromatic versions of the figure.

The Temporal Factor and Classification

Although there is still much to be learned about the cognitive activity involved in the classification procedure, it is clear that it begins very quickly after the figure enters the visual field. The first clue to this, generally disregarded by the Rorschach community, was noted by Stein (1949). He presented the blots tachistoscopically to two groups of subjects four times. The "ascending" group was shown the blots for one-hundredth of a second in the first administration, one-tenth of a second in the second administration, three seconds in the third, and given an unlimited time exposure in the fourth. The "descending" group (controls) received the four administrations with the exposure times reversed, that is, starting with the unlimited time exposure.

Unfortunately, the time intervals between testings were very brief and this rapid retest procedure probably confounded some of the results. Nonetheless, some clues about the rapidity of the response formulation may be gleaned from the results for the ascending group. In their first trial, with the blots exposed for only 10ms, the subjects averaged nearly 10 answers, with a range of five to 14 responses. When the blots were exposed for a full three seconds, two trials later, the average number of answers increased to 12, with a range

of eight to 17. However, Stein noted that, with the substantially longer exposure time, the frequencies of answers based exclusively or primarily on the contours of the blots increased markedly. In the third trial (three second exposure), a much greater number of Popular responses were given than had been the case when the blots were exposed for the brief intervals. In other words, once the blots were exposed long enough to permit a full scan of the field, considerable homogeneity occurred among the responses given.

In a related study, Horiuchi (1961) presented Cards III and VI tachistoscopically to groups of 80 nonpatients, 80 neurotics, and 80 schizophrenics for intervals of 0.10 seconds, 0.30 seconds, 1.0 second, and unlimited time. She found that 60 of the 80 nonpatients and nearly half of the neurotics and schizophrenics gave at least one response per blot with the exposure time of 100ms. When the exposure time was increased to 300ms, all of the nonpatients gave at least one response per blot, but the frequency of answers for the neurotic and schizophrenic groups did not increase. She also found that some persons from the neurotic and schizophrenic groups continued to have difficulty forming a differentiated response when the blots were exposed for a full second. She concluded that some of the mediational activity necessary to form an answer is inhibited in conditions of psychopathology.

Colligan and Exner (1985) used a tachistoscopic presentation of the blots with three groups of 36 subjects each, schizophreniform inpatients, orthopedic inpatients, and nonpatients. Each group was randomized into three subgroups of 12 persons each, and the blots were exposed to one subgroup for 200ms, the second for 400ms, and the third for 600ms. Subjects were instructed to give their responses after a tone, which was sounded 900ms after the exposure to allow sufficient time for the decay of the icon. They found that 62 of the 72 nonpsychiatric subjects were able to give at least one response to each blot, and nine of the 10 who gave refusals were in the subgroups with time exposures of 200ms or 400ms. Several of the nonpsychiatric subjects in each of the subgroups gave between 12 and 15 answers, but 17 of the 36 schizophreniform subjects gave fewer than 10 answers. Eight of those 17 were in the 600ms subgroup. These findings may support Horiuchi's postulate that mediation is impaired by psychopathology, but it might be more accurate to suggest that the schizophreniform individuals were much more defensive in the test situation.

Another finding from this study is probably more important. It concerns the appropriate use of the blot contours. Nearly 70% of all responses given by the nonpsychiatric subgroups included an appropriate use of contours. The answers in which the contours of the blot were violated occurred mainly to the more broken blots, II, III, VIII, and X, and many of those answers seem to reflect the principle of closure, that is, individuals created an imaginary line to encompass the parts of the blots, such as reporting that Card III is a face. This suggests that the exposure times were of insufficient length to permit a full scan of the more broken blots.

Colligan (1992) used the same design with 30 nonpatients, randomized into three groups of 10 each, and increased the exposure times. The figures were exposed to one group for 600ms, the second group for 800ms, and the third group for one second. Although the average *R* for each group is low, 11.2, 10.9, and 12.7 respectively, the range of *R* lengthened with exposure times. It extended from eight to 14 for persons in the 600ms group, eight to 16 in the 800ms group, and eight to 19 for those in the one second exposure group. The average number of Popular answers also increased with exposure times, 2.6, 3.2, and 4.7 respectively, as did the appropriate use of form. The *XA%* was 74% at 600ms, 83% at 800ms, and 86% at one second. Correspondingly, the *X–%* decreased from 26% at 600ms to 14% at one second. No minus answers were given to Card V by persons in any of the groups. It is the most solid of the figures and, apparently, easy to scan quickly. Interestingly, 50% of all of the minus answers were faces, most of which were given to Cards III, VIII, and X,

suggesting that even one second is probably not sufficient for a full scan of those broken figures.

The more broken the field, the more time will be necessary for the procedure of encoding and classifying, but extrapolating conservatively from the eye-tracking data and T-scope studies, it seems reasonable to assume that a period of two to three seconds after exposure is more than sufficient for encoding the field and classifying at least three, if not more, potential answers. Why then, do most people take at least twice that time, and often much longer, before delivering their first answer? It seems probable that the procedures of ranking and discarding are mainly responsible for the delay.

RANKING AND DISCARDING POTENTIAL ANSWERS

The instructions to people taking the test are deliberately brief, and no requirements or limits are implied concerning the number of responses. People who give only one answer to Card I are prompted for more, implying that only one response may not be sufficient. However, persons who give more than one response to Card I receive no confirmation that the yield is sufficient, except by the passive acceptance of the examiner. Some people seek direction by asking, "Is that enough" or "How many should I give," but unless a prompt is given on Card I, the response to such queries remains open-ended, *It's up to you.* This requires the person to decide how many of the potential answers that have been formulated to deliver.

The majority of persons confronted with this issue appear to be influenced by an economy principle in their decisions about the number of answers. To be sure, there are occasional compulsive-like people who seem prepared to give an infinite number of answers to each blot, but they are deterred from this with standard procedures. But most people are conservative with their yield of answers, and this probably accounts for the fact that most groups average about 22 responses. Whether the orientation to economize is a simple matter of efficiency, or a matter of defensive concealment is not certain, but probably both are involved. Practically all who take the test, even children, have some conception of psychological testing, and frequently will have some vague, often faulty, awareness of the inkblot test. People tend to think of testing in terms of educational models, which include right and wrong answers, high versus low scores, and passing or failing. Even when the examiner properly devotes time to an explanation of procedures and arranging for feedback, most people will still be influenced by prior sets and/or some apprehension about the situation. The desire to complete the test quickly and effectively is natural in such a situation, and probably contributes to some of the economy orientation.

Selecting by Rank Order

Another element that contributes to the economy orientation is the rank ordering process. As noted earlier, some blots or blot areas are easier to classify than others. Quite often a single blot or blot area will be identified more than once. For example, the whole Card I might be classified as a bat, a bird, and a butterfly. Some people elect to give two, or all three of those potential answers, but most persons will give only one of those responses. In doing so they discard the others. The selection of which of the three to deliver is based, at least in part, on some form of ranking, such as it looks more like a bat than either a bird or a butterfly.

This same process occurs whether the multiple potential answers are generated from the same blot area or from several areas. For instance, a person might classify the whole Card I as a bat and a butterfly, *D4* as a woman, *D2* as an animal, *D7* as a bird, and the *DdS29* areas as triangles. But that person is unlikely to give all six answers, and probably will give no more than two or three. At least in part, the decision of which two or three to give is influenced by a paired comparison rank ordering of the prospective answers. In a

pilot study, Martin and Thomas (1982) presented the blots twice, using a projector, to a class of 28 high school students. In the first presentation, each figure was exposed for one minute, and the students were instructed to *write* three responses per blot on a form provided. In the second exposure, the figures were shown again for 15 seconds, and the students were instructed to look at the figures again and select one of the three answers they had written to be *scored.* They were also to write a brief statement about why they selected that answer. Of the 280 statements, 159 could easily be categorized under the heading, "It looks most like that."[4]

Although it is clear that some form of paired comparison ranking, based on a similarity to known objects, does occur, it would be erroneous to assume that this procedure is *mainly* responsible for the selection of the answers that are delivered. There are at least three other elements that play a significant role in determining which of the potential answers will be delivered and which will be discarded.

Discarding through Censorship

As noted earlier, most people will approach psychological testing with some preconceptions. Some may concern psychological testing in general, and some may be specific to the Rorschach. Generally they result from bits and pieces of information concerning testing, but many of the sets with which a person approaches the test are derived from the more broadly based values. Whatever the origins, this composite of attitudes will often be very influential on the procedures of discarding and selecting. For instance, the MMPI *K* Scale has been demonstrated to have some relation to the orientation to make socially acceptable responses (Dahlstrom, Welsh, & Dahlstrom, 1972). In the Exner, Armbruster, and Mittman (1978) study cited earlier, the 20 persons scoring in the upper half of the MMPI *K* Scale distribution averaged 17 fewer responses than did those scoring in the lower half. This suggests that persons more oriented toward making acceptable responses tend to withhold or discard more responses than do persons not so oriented. These findings have provoked several other studies regarding the circumstances in which subjects may be more or less prone to withhold or discard responses.

In the study reported by Exner, Armbruster, and Mittman (1978), 10 therapists were asked to recruit two each of their own patients who had never taken the Rorschach. The patients were randomized so that each therapist tested one of his or her patients and the patient of another therapist with whom the examining therapist had no prior contact. The results show that those patients tested by their own therapist averaged 10 more responses to the test than did the controls, including significantly more sex responses (4.3 versus 0.8). Leura and Exner (1978) trained 10 junior high school teachers to administer the Rorschach and then used them in the same design. Each teacher was asked to recruit two volunteers from his or her seventh grade classes. The criterion for selection was that the students were progressing quite well in that class. The volunteer students were randomized, so that each teacher tested one of his or her own pupils, and a student from another school with whom the examining teacher had no prior contact. Students tested by their own teacher averaged nearly 16 more responses than did the controls.

The results of these studies seem to indicate that persons who feel emotionally and/or intellectually close to their examiner will deliver more and conceal less. However, this should not be translated to suggest that the well-trained examiner can or will have a significantly influential effect on how many responses are delivered. If the test is administered in the standardized manner, examiners can expect to obtain a normal distribution of responses. Exner (1974) demonstrated

[4] The comments for the total group varied considerably from "I don't know" to "I like those," but all categorized under the rubric of "It Looks Most Like That" ranged from form specified responses such as, "It has the wings and body like they are," to "It looks more like that than anything else there."

that some novice examiners may, because of their own difficulties and discomfort about learning the procedures of the test administration, cause fewer responses to be delivered than might ordinarily be the case. Apparently the discomfort of the examiner increases the apprehensiveness of the subject and leads to more withholding. But this is easily corrected with supervised experience in administration.

Goodman (1979) studied the effect of sex differences among examiners, using a design in which 10 male and 10 female examiners each tested two male and two female subjects. She also rated the examiners on an interpersonal "warmth" scale, using videotapes of them administering the TAT to a collaborator. Her results indicate no significant effects for the same-sex versus cross-sex pairings. She did find that the more experienced examiners, who also tended to be rated as warmer in their interactions, elicited a greater number of human responses than did the less experienced examiners. She also found that experienced examiners obtained more records of average length (17 to 27 responses) than did the examiners who were still in graduate school training. Although the element of rapport between the examiner and subject will contribute to censorship based discarding of potential answers, it seems likely that much censoring occurs because of sets about the test and value judgments concerning the acceptability of answers.

Exner and Leura (1976) used 30 male and 30 female volunteer nonpatients in a study concerning the ease with which objects can be perceived in the blots. None had prior exposure to the test. The 60 volunteers were randomized into two groups of 30 each, and seated at opposite ends of a hotel ballroom, each at their own small table. The groups were separated by a thick sliding wall. Each person was provided with a list of five answers per blot plus a location sheet on which all 50 answers were outlined. Each group of five answers was randomized for order, so that none appeared at the same point in the list for more than six persons in each group. Each list of five answers included a Popular response, two commonly given responses, and two others that are not given frequently. One of the five answers, the *target* response, had a content implying sexuality, injury, or violence.

For example, the five answers listed for Card I were, a bat, a mask, an animal, a bell, and the target answer, a naked woman. All of the target responses involve large or discrete areas of the blots. Subjects were instructed to look carefully at each blot as it was projected on a screen for an interval of 165 seconds, review the areas outlined on the location sheet, and decide the ease with which each could be recognized as compared with the remaining four answers. A rank of one was to be assigned for the answer easiest to see, two for the next easiest to see, and five for the one most difficult to see. The only difference between the two groups was that one was told the responses they would be ranking were among those given most often by normal people, whereas the second group was told that the responses represented those given most often by severely disturbed psychiatric patients.

When the groups were compared for their rankings, a significant difference occurred for 22 of the 50 answers, including eight of the 10 target responses. For example, only four of the 30 persons, told the responses came from disturbed patients, ranked the Card I target response (naked woman) 1 or 2, whereas 15 of the 30 assigned a rank of 5 to that answer. Conversely, 19 of the 30 subjects told that the answers were given frequently by normal individuals ranked the target answer 1 or 2. The responses listed for Card VI included an animal skin, a totem pole, a human profile, a dog, and the target response, a penis. Two people from the group told that the answers came from psychiatric patients ranked the target answer 2 and 3 while the remaining 28 ranked the target answer 4 or 5. In the group set to believe that the responses came from normal people, 19 ranked the target answer 1, and 4 others assigned the rank of 2. Only three individuals ranked the target answer 4 or 5.

Thomas, Exner, and Leura (1977) used another group of 60 nonpatients in a modification of this same design. Again, the group was randomized into

two groups of 30 each and seated at opposite ends of a ballroom separated by a sliding wall. They were provided with the same lists of five responses per blot used in the previous study, but with location sheets on which there were *no* outlines of the answers. These groups were given similar, but more distinctive sets concerning the origins of the responses. One group was told that the answers were given most commonly by very successful businessmen, while the second group was told that the responses were those given most often by inpatient schizophrenics. Instead of projecting the blots on a screen, each person was provided with a set of the Rorschach cards. The subjects were instructed to study each blot carefully, using as much time as might be necessary, and *find* each of the objects listed and, using a black marking pen, to outline each on the location sheet. After locating and outlining all five, they were to decide which was the easiest to see, the next easiest, and so on, using the same rankings of one to five that had been used in the earlier study. The results are similar to those of the earlier study.

The groups differed significantly for their rankings of 21 of the 50 responses, including nine of the 10 target responses. The 10 target responses and the frequencies for rankings are shown in Table 11.2. The ranks of 1 and 2, and 4 and 5 have been collapsed to represent the easier versus more difficult to see categories.

In each study, the difference in the rankings between the groups appears to be a function of the sets employed concerning the origins of the responses. It seems logical to postulate that the negative set (responses were given by psychiatric subjects) caused many to rank some responses as being more difficult to perceive, whereas those with the positive set felt less need to do so. Although studies of this kind are inconclusive about censoring operations that occur when selecting answers, they do appear to provide some clues about how the process may work. Even though the rank ordering operations might cause a potential answer to be ranked highly in terms of object similarity, the censoring operation might cause the response to be discarded because the person places a negative value judgment on it in light of the test situation.

STYLES AND TRAITS IN SELECTING RESPONSES

Although rank ordering and censoring seem to be important components of the response process, there are other elements that may supercede those operations in the discarding and selection

Table 11.2 Frequencies of Rankings for 10 Target Responses Given by Two Groups, with Ranks 1 and 2, and 4 and 5 Collapsed.

			Rankings					
			Group I Schizophrenia Set			Group 2 Businessmen Set		
Card	Area	Responses	1–2	3	4–5	1–2	3	4–5
I	*D4*	Naked woman	5	4	16*	18*	9	3
II	*D2*	Blood smears	1	10	19	14*	5	11
III	*D2*	Blood running down	5	11	14	16*	6	8
IV	*W*	Monster looming	21	6	3	20	9	1
V	*W*	Rams fighting	7	15	8	18*	10	2
VI	*D2*	Penis	3	9	18*	21*	5	4
VII	*D6*	Vagina	0	7	23*	13*	9	9
VIII	*W*	Open chest cavity	4	7	19	14*	5	11
IX	*D6*	Buttocks	0	12	18*	11*	11	8
X	*D9*	Blood stains	5	13	11	15*	7	8

*Statistically significant larger number of this rank. $p < .05$.

decisions. Among the more potent of these are the basic psychological characteristics of the individual. There is little doubt that they play a dominant role in determining which answers will be delivered. These are the features of people that cause them to be relatively consistent in many of their psychological operations and manifest behaviors. Historically, they have been identified as psychological habits, traits, styles, or dispositions. Whatever nomenclature is used, it represents the composite of the more dominant elements of the personality structure that breed behavioral preferences and create a tendency to redundancy in the selection of many coping responses.

These features are often reflected in the descriptions of a person rendered by those who know him or her well. For example, people may be described as being quiet, reserved, spontaneous, emotional, forceful, passive, and so on. Some people may be described as sturdy in the face of stress, while others may be thought of as easily disorganized under stressful conditions, and so on. If the person offering the description has frequent contact with the person being described, it is very likely that the description will be reasonably accurate. These characteristics are especially influential to many of a person's decision making operations when coping or problem solving are required. In that the task of the Rorschach requires a form of decision making, it is natural that the person will be influenced by these characteristics. In effect, the tendency of people to be behaviorally redundant manifests in Rorschach related operations, and creates a greater probability that certain classes of response will be selected for delivery than other classes of response that are also available.

The tendency to be redundant is the main feature causing the Rorschach to be temporally consistent. During the often stormy history of the Rorschach, those who disavowed its usefulness frequently pointed to problems of establishing satisfactory evidence of reliability. Unfortunately, most of the efforts to do so approached the issue by attempting to demonstrate that the test is internally consistent, using a split-half technique (Vernon, 1933; Hertz, 1934; Ford, 1946; Orange, 1953). Although the results were statistically significant, very few of the correlations fall into the .75 or higher range that would be required if a test is to be judged as truly internally consistent. The problem with this approach to reliability is the required assumption that the stimuli are equivalent, and will be equally likely to promote any class of response. The Rorschach figures are not equivalent stimuli. They differ for levels of complexity and are clearly different for the kinds of responses they are likely to generate.

In developing the Comprehensive System, the issue of reliability has been addressed by studying the temporal consistency of classes of response. The operational hypothesis posed is that people do have preferred response styles that manifest in the majority of their responses, and evidence for those styles should be consistently evident in repeated testings. Holzberg (1960) has questioned the usefulness of the test-retest model for the Rorschach, arguing that personality variables might not be consistent over time, and that a different set might occur during the second testing because of the memory input created by the earlier testing. Neither of these objections is very convincing. A host of data exist indicating that many so-called personality traits do remain consistent over time (London & Exner, 1978), and the findings of the Exner, Armbruster, and Mittman (1978) study suggest that if memory is an important variable, it is less contingent on recalling what was seen, and more on recalling what was reported.

Numerous temporal consistency studies have been completed, using groups of adults and children, patients and nonpatients. The intervals between first and second tests have varied from a few days to many months. When considered with regard to a consistency of response styles or dispositions, two long term retest studies offer the most substantial data. In the first (Exner, Armbruster, & Viglione, 1978), 100 nonpatient adults, 50 male and 50 female, were retested after 36 to 39 months. In the second (Exner, Thomas, & Cohen,

1983), 50 nonpatient adults, 25 male and 25 female, were retested after 12 to 14 months. Some additional calculations from the 1983 data were made after the protocols were rescored for new variables (Exner, 1999). Retest correlations from these studies are shown in Table 11.3.

An examination of the data for the 50 persons retested after approximately one year reveals that four of the correlations exceed .90, and 25 others fall between .81 and .89. Ten of the correlations fall below .75, a finding that is not surprising, because the psychological characteristics related to all 10 can be influenced by *state* as well as *trait* features. The data for the retests taken at approximately three years are quite similar. One of the correlations falls at .90, and 18 others are between .80 and .87. Six, all related to state conditions, fall below .70. The psychometric purist might argue, with some merit, that correlations falling below .80, or even below .85, are not sufficient to support a claim of stability or consistency for a variable. This argument is most relevant to those variables having retest correlations between .70 and .79. Seven of these appear in the data for the one year retest, and 10 in the data for the three year retest. The features and/or operations related to these variables are not as consistent as the features and operations related to the variables with higher retest correlations, yet each accounts for at least one-half of the variance. Thus, it seems reasonable to postulate that these variables relate to characteristics having some considerable stability, but which are also more subject to influence by other conditions, including lengthy time intervals.

It is also important to note that none of the 26 single variables listed in Table 11.3, *taken alone,* have a critical impact on the interpretation of the test. The 15 ratios and percentages shown in Table 11.3 have a much more important role in interpretation, and most of those do have retest correlations of .80 or greater. Those having correlations of less than .80 are much more related to state influences.

Whereas the data in Table 11.3 indicate that most of the characteristics represented by Rorschach scores are very stable among nonpatient adults over long periods, the same is not true for children. Exner and Weiner (1982) have reported relatively low retest correlations for six-year-olds who were retested at age eight, and for nine-year-olds retested when they were 12. The results of an eight-year longitudinal study of 57 nonpatient youngsters, tested first at age eight and retested at each two year interval through the age of 16, show that retest correlations for most variables tend to remain low until the interval between 14 and 16 years of age (Exner, Thomas, & Mason, 1985). However, this should not be interpreted to mean that the traits or styles of the child are not stable over briefer intervals, or that those features will have little influence on the selection of responses. Retest correlations for most variables are quite respectable, for both children and adults, when the second test is taken in less than one month. Table 11.4 includes the results of three such studies. One concerns 25 nonpatient eight-year-olds retested after seven days (Exner & Weiner, 1982). In the other two, involving 35 nonpatient nine-year-olds and 35 nonpatient adults, the retest was administered after approximately three weeks (Thomas, Alinsky, & Exner, 1982).

The correlations for the eight-year-olds retested after seven days include nine which are at .90 or greater and 13 others between .80 and .89. Those for the nine-year-olds include eight at .90 or greater, and 17 others falling between .80 and .89. The correlations for the adults include six at .90 or greater and 25 between .80 and .89. The number of high correlations yielded from these brief interval retest studies tends to favor the postulate that these characteristics do play an important role in the selection of answers to be delivered. It could be argued, following from Holzberg (1960), that the memory factor may play an important role when brief interval retesting is done, thereby causing some correlations to be spuriously elevated. But there is evidence to indicate that this is not the case.

Exner (1980) recruited 60 nonpatient eight-year-olds from four elementary schools under the guise of needing *practice* subjects for examiners in training. Thus, each child volunteered for the

Table 11.3 Correlation Coefficients for Nonpatient Groups of 50 Adults Retested after 12 to 14 Months and 100 Adults Retested after 36 to 39 Months.

Variable	Description	1-Year Retest *r*	3-Year Retest *r*
R	Number of Responses	.86	.79
P	Popular responses	.83	.73
Zf	Z Frequency	.85	.83
F	Pure Form	.74	.70
M	Human Movement	.84	.87
FM	Animal Movement	.77	.72
m	Inanimate Movement	.26	.39
a	Active Movement	.83	.86
p	Passive Movement	.72	.75
FC	Form Color Responses	.86	.86
CF	Color Form Responses	.58	.66
C	Pure Color	.56	.51
CF+C	Color Dominant Responses	.81	.79
SumC	Sum Weighted Color	.82	.86
SumT	Texture Responses	.91	.87
SumC'	Achromatic Color Responses	.73	.67
SumY	Diffuse Shading Responses	.31	.23
SumV	Vista Responses	.87	.81
FD	Dimension Responses	.88	.83
Fr+rF	Reflection Responses	.82	.78
(2)	Pair Responses	.81	.83
DV+DR	Deviant Responses	.72	.79**
INC+FAB	Inappropriate Combinations	.89	.82
COP	Cooperative Movement	.81	*
AG	Aggressive Movement	.82	*
MOR	Morbid Content	.71	*
Ratios and Percentages			
L	Lambda	.78	.82
EA	Experience Actual	.83	.85
es	Experienced Stimulation	.64	.72
Adjes	Adjusted es	.82	*
D	Stress Tolerance Index	.91	.83
XA%	Appropriate Form Use	.89	*
WDA%	Appropriate Form Use To W+D	.92	*
X+%	Appropriate/Common Good Form	.86	.80
X–%	Distorted Form Use	.92	.86
Afr	Affective Ratio	.82	.90
3r+(2)/R	Egocentricity Index	.89	.87
WSum6	Weighted Sum Special Scores	.86	*
Blends	Multiple Determinants	.62	.67
Intell	Intellectualization Index	.84	*
Isolate/R	Isolation Index	.84	*

* = Not coded or calculated in the Comprehensive System at the time of the study.
**= Includes only DV as DR not coded in the Comprehensive System at the time of the study.

Table 11.4 Correlation Coefficients for 25 Nonpatient 8-Year-Olds Retested after 7 Days, 35 Nonpatient 9-Year-Olds, and 35 Nonpatient Adults Retested after Approximately 3 Weeks.

		Eight-Year-Olds 7-Day Retest	Nine-Year-Olds 3-Week Retest	Adults 3-Week Retest
Variable	Description	*r*	*r*	*r*
R	No. of Responses	.88	.87	.84
P	Popular responses	.86	.89	.81
Zf	Z Frequency	.91	.92	.89
F	Pure Form	.79	.80	.76
M	Human Movement	.90	.87	.83
FM	Animal Movement	.75	.78	.72
m	Inanimate Movement	.49	.20	.34
a	Active Movement	.91	.91	.87
p	Passive Movement	.86	.88	.85
FC	Form Color Responses	.90	.84	.92
CF	Color Form Responses	.76	.74	.68
C	Pure Color	.72	.64	.59
CF+C	Color Dominant Responses	.89	.92	.83
Sum C	Sum Weighted Color	.88	.87	.83
SumT	Texture Responses	.86	.92	.96
SumC'	Achromatic Color Responses	.77	.74	.67
SumY	Diffuse Shading Responses	.42	.17	.41
SumV	Vista Responses	.96	.93	.89
FD	Dimension Responses	.74	.81	.90
Fr+rF	Reflection Responses	.83	.80	.89
(2)	Pair Responses	.74	.77	.83
DV+DR	Deviant Responses	.71**	.76**	.74
INC+FAB	Inappropriate Combinations	.72	.81	.92
COP	Cooperative Movement	*	*	*
AG	Aggressive Movement	*	*	.81
MOR	Morbid Content	*	*	.83
Ratios and Percentages				
L	Lambda	.82	.84	.76
EA	Experience Actual	.85	.87	.84
es	Experienced Stimulation	.74	.70	.59
Adjes	Adjusted es	*	*	.79
D	Stress Tolerance Index	.93	.91	.88
XA%	Appropriate Form Use	*	*	*
WDA%	Appropriate Form Use to W+D	*	*	*
X+%	Extended Good Form	.95	.92	.87
X–%	Distorted Form Use	.83	.80	.88
Afr	Affective Ratio	.91	.91	.85
3r+(2)/R	Egocentricity Index	.94	.86	.90
WSum6	Weighted Sum Special Scores	*	*	*
Blends	Multiple Determinants	.57	.64	.71
Intell	Intellectualization Index	*	*	*
Isolate/R	Isolation Index	*	*	*

* = Not coded or calculated in the Comprehensive System at the time of the study.
** = Includes only DV as DR not coded in the Comprehensive System at the time of the study.

study by his or her parents, anticipated being tested at least twice, during regular school hours, in a five- to seven-day interval. Ten very experienced examiners were used. None had any awareness of the nature of the study, assuming that they were collecting data in a routine reliability investigation. When the first test was administered, the project director accompanied the child from the classroom to the testing room provided by the school and introduced the examiner. The second test was administered three or four days later. The 60 children had been randomized previously into two groups of 30 each, and the procedure for the control subjects was the same for the second test as the first. This procedure was altered for the subjects in the experimental group. While accompanying them to the testing room, the project director would pause and ask the child to assist in the solution of an important problem concerning the training of the examiners. The "problem," as outlined quickly by the project director, was that the "trainees" were hearing the same answers repeatedly, and that the training would be enhanced considerably if the child would "try hard" to remember the answers that he or she gave in the first test, and promise not to repeat any of those in the test to be taken. In return, the project director offered a reward of 50 cents. All of the children in the experimental group promised to change their answers.

The retest correlations of the two groups are almost identical, with two exceptions. In the retest, the children in the experimental group gave significantly fewer Pure Form responses than in the first test and significantly more responses that included the achromatic or shading features of the blots. Otherwise, the correlations are essentially no different for the two groups, and very similar to those shown in Table 11.3 for the eight-year-olds retested after only seven days. A critical issue is whether the youngsters in the experimental group actually gave different answers. This was addressed by randomizing the 60 pairs of records into three groups of 20 pairs each and assigning each group to one of three judges who were told that the records were from a reliability study. The judges were instructed to read each pair of records and check all responses in the second test that were the same (or nearly the same) as those in the first test. The comparisons revealed that 481 of the 546 responses (86%) given in the second test by the control subjects were replications or near replications of answers that they had given in the first test. On the other hand, only 77 of the 551 answers (14%) given by experimental subjects in the second test were the same or similar to the first test answers. Thus, even though the youngsters in the experimental group generally did honor the promise to give different responses in the second test, the answers that they tended to select show a distribution of scores, subject by subject, similar to those of the first test.

There are other structural data, not shown in Tables 11.2 or 11.3, which are also important to the study of the consistency hypothesis as related to the selection of answers. Among these are three relationships, or ratio directions. Each is important to the interpretation of the test, either because of the directionality shown in the relationship (which side of the ratio is greater) and/or the magnitude of the difference. The first, the *Erlebnistypus* (*EB*) involves the relationship of *M* to *SumC*. Both direction and magnitude are important to the interpretation of this relationship. In the one-year retest study of 50 adult nonpatients cited earlier, the first tests of 40 subjects showed an *EB* in which one side of the ratio was two or more points greater than the other side. In the second test, 39 of those 40 persons continued to show at least a two-point difference in the *EB*, and only one had changed for the direction of the difference (Exner, 1999). In the study involving the retest of 100 nonpatient adults after three years, 83 of the first test records showed at least a two-point difference in the *EB*. In the retest, 77 of those 83 continued to have the difference of two points or more and only two had changed directionality.

A second important ratio concerns the relationship between *FC* and *CF* + *C*. In the one-year retest study, 27 of the 50 participants had

FC values that were equal to, or greater than, the value for *CF + C*. The remaining 23 had higher values for *CF + C*. The retest data show that 26 of the 27 persons having higher *FC* values continued to show that directionality (Exner, 1999). In the three-year retest study, 57 of the 100 first records had a difference of two points or more for this ratio. In the retest, 50 of the 57 continued to have the two or more point difference and none had changed directionality.

The third is the ratio of active to passive movement. In this relation, more active movement is always expected, but the magnitude of the difference is also important. In the one-year retest study, 47 of the 50 first test records had values for active movement that were equal to, or greater than, the values for passive movement. The retest protocols show that 46 of 47 continue to have values for active movement equal to, or greater than passive movement (Exner, 1999). In the three-year retest, a two point or more difference was noted in 76 of the first records, with 60 being higher for active movement. In the retest, 57 of those 60 were still two or more points higher on the active side, and 11 of the original 16 high passive records remained at least two points higher in that direction.

The composite of data regarding the consistency of Rorschach scores and ratios, during both lengthy and brief intervals, and even under conditions where different responses are prompted, affords substantial support for the proposition that the traits or styles of the person will be quite influential in the selection of which potential answers to deliver. Yet, there is still another element that will often play a crucial role in the final decisions about which responses to deliver.

PSYCHOLOGICAL STATES AND RESPONSE SELECTION

The consistency of many personality characteristics and their consequent behaviors is contingent upon some relative consistency in the conditions that evoke them. In other words, the habits, traits, or response styles of a person are probabilistic. Certain classes of behaviors are likely to occur under certain classes of conditions. The conditions include *both internal and external elements*. For instance, a person who is an avid outdoor enthusiast may be far less likely to engage in outdoor activities if temperatures fall well below zero or above 110°. Sedentary activity may replace some of the common activities. Likewise, a person is less likely to engage in outdoor activities if his or her temperature exceeds 101° or if an irritating stomach upset occurs. The person has not changed, but external or internal conditions have changed in a way to cause other behaviors to be evoked.

Similarly, alterations in routine behavior, may result from changes in the psychological state of the individual. Increases or decreases in needs and/or emotions, or the unexpected experience of stress, or the onset of various psychopathological states can have the effect of evoking new behaviors that add to, or replace, preexisting orientations. In most instances, the basics of the person do not change, but some unexpected behaviors may occur. Sometimes, there is only a slight shift in behavior but, in other instances, the deviations from expected behaviors may be much more marked. These are *state* phenomena, which tend to supersede the routine psychological functioning of the person, or, at least, will stimulate the addition of behaviors that are not ordinarily part of the psychological routine.

As has been emphasized, Rorschach responses can be viewed as a sampling of decision making behavior. Thus, if a psychological state exists that will alter, or add to the decision making routines of a person, that state can also influence the selection of responses to be given during the test. For example, two of the variables listed in Tables 11.2 and 11.3 have low retest reliabilities, for both lengthy and brief retest intervals. These are the *m* and *SumY* variables, both of which are expected to occur about once in a record. Both relate to situational stress (Shalit, 1965; Armbruster, Miller, & Exner, 1974; Exner, Armbruster, Walker, & Cooper, 1975;

Exner, 1978; Exner & Weiner, 1982). Thus, if the frequencies for either or both are elevated, it signals that the presence of some situationally related phenomenon which is stimulating the mental and/or emotional experiences that occur when concerns about loss of control and feelings of helplessness or paralysis exist.

Another illustration of how the selection of Rorschach answers may be influenced by a state condition is exemplified by the texture variable. Approximately 70% to 80% of all nonpatient adults give one texture answer and, for this reason, the retest reliabilities for *SumT* usually ranges from .80 to .90. If a person has recently experienced a significant emotional loss, the average number of texture responses will often increase. Exner and Bryant (1974) found an average of nearly four texture responses in the records of 30 adults who had recently separated from close emotional relationships. When retested, approximately 10 months later, 21 of the 30 reported that they had established new relationships or reconstructed the one that had been fractured, and they uniformly gave substantially fewer texture responses in the second test than in the first. The nine people who reported a continuing sense of loss all gave three or more texture responses in the second test.

Although many psychological states are transient, others can form a more enduring overlay to the primary personality structure. Many psychopathological states have this impact, and just as they influence a broad spectrum of behaviors, they are influential in the selection of Rorschach answers. Severe and/or enduring depression is an example of this. Depressed people tend to give higher frequencies of vista and achromatic color answers than do other groups. They also tend to give more MOR responses, and usually have low values on the Egocentricity Index. Haller and Exner (1985) used a design similar to that employed by Exner (1980) with children, with 50 patients, admitted with presenting symptoms of depression and/or helplessness. All were retested three or four days after the first test, and a randomly selected half were asked to give different answers in the second test. That group repeated about 33% of their first test answers, whereas the control subjects repeated nearly 70% of their first test responses. In spite of the fact that the experimental group gave more than 68% new answers in the second test, the retest reliabilities for both groups were very similar to those reported by Exner for children in the same design. Moreover, the group did not differ significantly from one to another for any of the variables related to depression.

As the state endures, so does its influence, and the more severe the state may be, its impact on decision making will be greater. Exner et al. (1985) found that the retest reliabilities for the major indices of depression remained very high among inpatient adolescents diagnosed as having a major depressive disturbance even after having been in treatment for nearly one year. On the other hand, as the state dissipates, so too does its influence on the selection of Rorschach answers. Exner, Cohen, and Hillman (1984) retested 46 persons, diagnosed by *DSM-III* criteria as having a major depressive disorder, at the termination of their treatment. All had begun treatment as inpatients and continued treatment as outpatients for a period averaging almost two years. The retest correlations for each of the variables related to depression were very low, ranging from .19 for the vista variable, to .33 for the presence of MOR responses.

In effect, the psychological state of the person taking the Rorschach will contribute in the final selection of the answers that are delivered. The influence of the state in the selection of responses will probably be proportionate to its impact on the individual, and its duration. It will not always supersede other influencing traits, styles, or habits, but it may.

PROJECTION AND THE RESPONSE PROCESS

Does projection occur in the selection and/or discarding of all responses? That position can be argued only if the definition of projection is

extended to include all decision operations, but that is a simplistic and overinclusive position. If the selection of a response is based exclusively on the classification and rank ordering operations, it seems highly unlikely that projection is involved. The more common responses that are formed mainly by using contours, such as a bat, two dogs, a butterfly, a tree, and so on, are the best examples of answers for which there is no evidence of projection. Similarly, much or all of the discarding of potential answers, as a result of rank ordering or censorship, probably is not influenced much by the projective process. Most discarding of this nature is prompted more by attitudes and values of the person, and his or her perception of the test situation.

It is also difficult to reason that the traits or styles of the person are significantly impacted by projection as such. To the contrary, these are enduring characteristics that will be influential in directing the projective process when it does occur, and the same relationship exists between state influences and projection. When a state is intense and broadly encompassing, it can easily give rise to much rich projective material.

Projections can and do occur in the Rorschach and their presence often adds considerably to the interpretive yield of the test, but it is important to discriminate projected material from that which is not. Exner (1989) presented data suggesting that there are probably two types of projected responses. One type is formed during the Phase I operations and the second is formed during the Phase II or Phase III operations.

Projection and Phase I

The first type of projection that may occur in Rorschach answers apparently provokes some sort of distortion and/or misperception during the input and classification process. For instance, most people who are shown a ball, and asked what it is, will usually respond that it is a ball. The distal features of the object reduce the parameters of its definition rather sharply. It can be identified functionally ("Something you throw"), or even syncretistically ("It's a thing that is made by humans"), but the range of appropriate responses is limited.

If a person who is not perceptually impaired misidentifies the ball by calling it an airplane, or a devil, or a kidney, it is not unreasonable to assume that some element of projection has occurred because the distal field has been markedly distorted and/or ignored. Even though distal elements in figure tend to facilitate certain responses or classes of response, thereby reducing the likelihood that projection will be involved in Phase I operations, classifications of blots or blot areas do occur in ways that violate or ignore those features. Technically, they are minus answers, and if they are not the product of some neurophysiologically related perceptual dysfunction, it seems logical to postulate that they are the result of a form of mediation in which internal psychological sets or operations have superseded a reality oriented translation of the field. Thus, a form of projection is involved.

Projection and Phases II and III

Although it is possible for projection to play a role during the Phase I classifications, it is much more likely that the impact of the projection, if it occurs, will be greater during the Phase II and III operations. During these operations, a type of imaginative projection is manifest when the person overelaborates on, or departs from, the stimulus field. For instance, most people who are asked to multiply two times two will answer four, but the answer could be embellished in a projected manner such as, *The answer is four, which is my favorite number because it represents the four seasons of the year that are so important to the cycle of life.*

An elaboration such as this reflects something about the person who gave it because there is nothing in the question posed that provokes it. In this example, the nature of the task, that is, multiplying numbers, reduces the probability that projection will occur. It can occur, but usually will not. In the

Rorschach, the parameters of both the task and the field are broader. While the limited ambiguity of the field and the nature of the task do not encourage projection, they do not prohibit or discourage the unique translations or embellishments that almost certainly have some projected properties. As a result, projected features are included when some answers are formed, and they ultimately appear in the response that is given.

The majority of answers that include this kind of projection do not require much interpretive translation because the embellishment is usually obvious by its departure from, or overelaboration of the field. In these responses, the object(s) is described in a specific or unique manner. Many such projections occur in movement answers, but many others occur in responses in which there is no movement, but in which the person embellishes his or her description of the object considerably. When they occur, these types of projected answers are very direct reflections concerning the feelings and/or behaviors of people.

During interpretation, it is important to recognize that not all of the responses, or all of the verbiage, represent projected material. Many records will contain more responses that contain no projections than those that do have that feature. In fact, it is uncommon, but not infrequent, to find records in which no projected material appears to exist. Usually these are closely guarded, relatively short protocols in which each response is limited to a few words. Those records are no less valid than the more elaborated protocols, but they lack the rich, idiosyncratic features of the person that appear in projected responses, and which can contribute much in fleshing out a description of the person.

SUMMARY

The nature of the Rorschach task provokes a complicated process, that includes processing, classification, conceptualization, decision making, and lays open the psychological door for projection to occur. When the reasonably simple standard procedures for administering the test are faithfully employed, this process yields substantial information about habits, traits, and styles, the presence of states, and about many other variables that can be listed under the broad rubric encompassed by the term personality. While the process is complex, the test itself is not a complicated assessment tool for the adept user, especially when the user understands the nature of the test and how it works. However, the procedures for administering the test are also delicate. If they are violated by intent, or by the naivete of the user, the process of the test may be altered substantially, and the method can become reduced from a test to a composite of verbiage that, at best, will be of questionable interpretive value.

REFERENCES

Abramson, L. S. (1951). The influence of set for area on the Rorschach test results. *Journal of Consulting Psychology, 15,* 337–342.

Ames, L. B., Learned, J., Metraux, R. W., & Walker, R. N. (1952). *Child Rorschach responses.* New York: Hoeber-Harper.

Ames, L. B., Metraux, R. W., & Walker, R. N. (1971). *Adolescent Rorschach responses.* New York: Brunner/Mazel.

Armbruster, G. L., Miller, A. S., & Exner, J. E. (1974). *Rorschach responses of parachute trainees at the beginning of training and prior to the first jump.* Rorschach Workshops (Study No. 201, unpublished).

Attneave, F. (1954). Some information aspects of visual perception. *Psychological Review, 61,* 183–193.

Beck, S. J. (1945). *Rorschach's test II: A variety of personality pictures.* New York: Grune & Stratton.

Beck, S. J., Beck, A. G., Levitt, E. E., & Molish, H. B. (1961). *Rorschach's test I: Basic processes* (3rd ed.). New York: Grune & Stratton.

Cattell, R. B. (1951). Principles of design in "projective" or misperceptive tests of personality. In H. Anderson & G. Anderson (Eds.), *Projective techniques.* Englewood Cliffs, NJ: Prentice-Hall.

Coffin, T. E. (1941). Some conditions of suggestion and suggestibility: A study of certain attitudinal and situational factors in the process of suggestion. *Psychological Monographs, 53*(Whole No. 241).

Colligan, S. C. (1992). *Responses of nonpatients to a tachistoscopic presentation of the Rorschach.* Washington, DC: Society for Personality Assessment.

Colligan, S. C., & Exner, J. E. (1985). Responses of schizophrenics and nonpatients to a tachistoscopic presentation of the Rorschach. *Journal of Personality Assessment, 49,* 129–136.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook* (Vol. 1., rev.). Minneapolis: University of Minnesota Press.

Dinoff, M. (1960). Subject awareness of examiner influence in a testing situation. *Journal of Consulting Psychology, 24,* 465.

Dubrovner, R. J., VonLackum, W. J., & Jost, H. (1950). A study of the effect of color on productivity and reaction time in the Rorschach test. *Journal of Clinical Psychology, 6,* 331–336.

Erdberg, S. P., & Shaffer, T. W. (1999). *International symposium on Rorschach nonpatient data: Findings from around the world.* Amsterdam: XVI International Congress of Rorschach and Projective Methods.

Exner, J. E. (1959). The influence of chromatic and achromatic color in the Rorschach. *Journal of Projective Techniques, 23,* 418–425.

Exner, J. E. (1961). Achromatic color in Cards IV and VI of the Rorschach. *Journal of Projective Techniques, 25,* 38–40.

Exner, J. E. (1974). *The Rorschach: A Comprehensive System. Volume I.* New York: Wiley.

Exner, J. E. (1978). *The Rorschach: A Comprehensive System. Volume 2. Recent research and advanced interpretation.* New York: Wiley.

Exner, J. E. (1980). But it's only an inkblot. *Journal of Personality Assessment, 44,* 562–577.

Exner, J. E. (1983). Rorschach assessment. In I. B. Weiner (Ed.), *Clinical methods in psychology.* New York: Wiley.

Exner, J. E. (1989). Searching for projection in the Rorschach. *Journal of Personality Assessment, 53,* 520–536.

Exner, J. E. (1996). Critical bits and the Rorschach response process. *Journal of Personality Assessment, 67,* 464–477.

Exner, J. E. (1999). The Rorschach: Measurement concepts and issues of validity. In S. B. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement.* Mahwah, NJ: Erlbaum.

Exner, J. E., & Armbruster, G. L. (1974). *Increasing R by altering instructions and creating a time set.* Rorschach Workshops (Study No. 209, unpublished).

Exner, J. E., Armbruster, G. L., & Mittman, B. (1978). The Rorschach response process. *Journal of Personality Assessment, 42,* 27–38.

Exner, J. E., Armbruster, G. L., & Viglione, D. (1978). The temporal stability of some Rorschach features. *Journal of Personality Assessment, 42,* 474–482.

Exner, J. E., Armbruster, G. L., Walker, E. J., & Cooper, W. H. (1975). *Anticipation of elective surgery as manifest in Rorschach records.* Rorschach Workshops (Study No. 213, unpublished).

Exner, J. E., & Bryant, E. L. (1974). *Rorschach responses of subjects recently divorced or separated.* Rorschach Workshops (Study No. 206, unpublished).

Exner, J. E., Cohen, J. B., & Hillman, L. B. (1984). *A retest of 46 major depressive disorder patients at the termination of treatment.* Rorschach Workshops (Study No. 275, unpublished).

Exner, J. E., & Leura, A. V. (1976). *Variations in ranking Rorschach responses as a function of situational set.* Rorschach Workshops (Study No. 221, unpublished).

Exner, J. E., Martin, L. S., & Cohen, J. B. (1983). *Card by card response frequencies for patient and nonpatient populations.* Rorschach Workshops (Study No. 276, unpublished).

Exner, J. E., Thomas, E. A., & Cohen, J. B. (1983). *The temporal consistency of test variables for 50 nonpatient adults after 12 to 14 months.* Rorschach Workshops (Study No. 281, unpublished).

Exner, J. E., Thomas, E. A., & Mason, B. (1985). Children's Rorschachs: Description and prediction. *Journal of Personality Assessment, 49,* 13–20.

Exner, J. E., & Weiner, I. B. (1982). *The Rorschach: A Comprehensive System. Volume 3: Assessment of children and adolescents.* New York: Wiley.

Exner, J. E., & Weiner, I. B. (1995). *The Rorschach: A Comprehensive System. Volume 3: Assessment of children and adolescents* (2nd ed.). New York: Wiley.

Exner, J. E., & Wylie, J. R. (1976). *Alterations in frequency of response and color articulation as related to alterations in the coloring of specific blot areas.* Rorschach Workshops (Study No. 219, unpublished).

Exner, J. E., & Wylie, J. R. (1977). *Differences in the frequency of responses to the D1 area of Card X using an achromatic version.* Rorschach Workshops (Study No. 237, unpublished).

Fisher, D. F., Monty, R. A., & Senders, J. W. (Eds.). (1981). *Eye movements: Cognition and visual perception.* Hillsdale, NJ: Erlbaum.

Ford, M. (1946). *The Application of the Rorschach test to young children.* Duluth: University of Minnesota, Institute of Child Welfare.

Frank, L. K. (1939). Projective methods for the study of personality. *Journal of Psychology, 8,* 389–413.

Gibby, R. G. (1951). The stability of certain Rorschach variables under conditions of experimentally induced sets: I. The intellectual variables. *Journal of Projective Techniques, 15,* 3–26.

Goetcheus, G. (1967). *The effects of instructions and examiners on the Rorschach.* Unpublished master's thesis, Bowling Green State University, Bowling Green, OH.

Goodman, N. L. (1979). *Examiner influence on the Rorschach: The effect of sex, sex-pairing and warmth on the testing atmosphere.* Doctoral dissertation, Long Island University, Brooklyn, NY.

Gross, L. (1959). Effects of verbal and nonverbal reinforcement on the Rorschach. *Journal of Consulting Psychology, 23,* 66–68.

Haller, N., & Exner, J. E. (1985). The reliability of Rorschach variables for inpatients presenting symptoms of depression and/or helplessness. *Journal of Personality Assessment, 49,* 516–521.

Halpern, F. (1953). *A clinical approach to children's Rorschachs.* New York: Grune & Stratton.

Hersen, M., & Greaves, S. T. (1971). Rorschach productivity as related to verbal reinforcement. *Journal of Personality Assessment, 35,* 436–441.

Hertz, M. R. (1934). The reliability of the Rorschach ink-blot test. *Journal of Applied Psychology, 18,* 461–477.

Hochberg, J. (1981). Levels of perceptual organization. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization.* Hillsdale, NJ: Erlbaum.

Holzberg, J. D. (1960). Reliability re-examined. In M. Rickers-Ovsiankina (Ed.), *Rorschach psychology.* New York: Wiley.

Horiuchi, H. (1961). A study of perceptual process of Rorschach cards by tachistoscopic method on movement and shading responses. *Journal of Projective Techniques, 25,* 44–53.

Hutt, M., Gibby, R. G., Milton, E. O., & Pottharst, K. (1950). The effect of varied experimental "sets" upon Rorschach test performance. *Journal of Projective Techniques, 14,* 181–187.

Klopfer, B., & Kelley, D. M. (1942). *The Rorschach technique.* Yonkers-on-Hudson, NY: World Books.

Leura, A. V., & Exner, J. E. (1978). *Structural differences in the records of adolescents as a function of being tested by one's own teacher.* Rorschach Workshops (Study No. 265, unpublished).

London, H., & Exner, J. E. (1978). *Dimensions of personality.* New York: Wiley.

Magnussen, M. G. (1960). Verbal and nonverbal reinforcers in the Rorschach situation. *Journal of Clinical Psychology, 16,* 167–169.

Martin, L. S., & Thomas, E. E. (1982). *Selection of preferred responses by high school students.* Rorschach Workshops (Study No. 278, unpublished).

Matarazzo, J. D., & Mensh, I. N. (1952). Reaction time characteristics of the Rorschach test. *Journal of Consulting Psychology, 16,* 132–139.

Meer, B. (1955). The relative difficulty of the Rorschach cards. *Journal of Projective Techniques, 19,* 43–53.

Meer, B., & Singer, J. L. (1950). A note on the "father" and "mother" cards in the Rorschach inkblots. *Journal of Consulting Psychology, 14,* 482–484.

Miale, F. R., & Harrower-Erikson, M. R. (1940). Personality structure in the psychoneuroses. *Rorschach Research Exchange, 4,* 71–74.

Murray, H. A. (1938). *Explorations in personality.* New York: Oxford University Press.

Neisser, U. (1976). *Cognition and reality.* New York: Appleton-Century-Crofts.

Orange, A. (1953). Perceptual consistency as measured by the Rorschach. *Journal of Projective Techniques, 17,* 224–228.

Pascal, G., Ruesch, H., Devine, D., & Suttell, B. (1950). A study of genital symbols on the Rorschach test: Presentation of method and results. *Journal of Abnormal and Social Psychology, 45,* 285–289.

Phillips, L., & Smith, J. G. (1953). *Rorschach interpretation: Advanced technique.* New York: Grune & Stratton.

Piotrowski, Z. (1957). *Perceptanalysis.* New York: Macmillan.

Pomerantz, J. R., & Kubovy, M. (1981). Perceptual organization: An overview. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization.* Hillsdale, NJ: Erlbaum.

Rabin, A. I., & Sanderson, M. H. (1947). An experimental inquiry into some Rorschach procedures. *Journal of Clinical Psychology, 3,* 216–225.

Shalit, B. (1965). Effects of environmental stimulation on the M, FM, and m responses in the Rorschach. *Journal of Projective Techniques and Personality Assessment, 29,* 228–231.

Silva, D. (2002). The effect of color on the productivity in Card X of the Rorschach. *Rorschachiana, 25,* 123–138.

Stein, M. I. (1949). Personality factors involved in the temporal development of Rorschach responses. *Rorschach Research Exchange, 13,* 355–414.

Thomas, E. A., Alinsky, D., & Exner, J. E. (1982). *The stability of some Rorschach variables in 9-year-olds as compared with nonpatient adults.* Rorschach Workshop (Study No. 441, unpublished).

Thomas, E. A., Exner, J. E., & Leura, A. V. (1977). *Differences in ranking responses by two groups of nonpatient adults as a function of set concerning the origins of the responses.* Rorschach Workshop (Study No. 251, unpublished).

Vernon, P. E. (1933). The Rorschach inkblot test. II. *British Journal of Medical Psychology, 13,* 179–205.

Wickes, T. A. (1956). Examiner influence in a testing situation. *Journal of Consulting Psychology, 20,* 23–26.

4.18
Therapeutic Assessment: Linking Assessment and Treatment

MARK E. MARUISH
Strategic Advantage, Minneapolis, MN, USA

4.18.1 INTRODUCTION 526

4.18.2 THE CURRENT PRACTICE OF PSYCHOLOGICAL ASSESSMENT IN THE THERAPEUTIC ENVIRONMENT 528

4.18.3 PSYCHOLOGICAL ASSESSMENT AS A THERAPEUTIC ADJUNCT 530
4.18.3.1 Psychological Assessment for Clinical Decision-making 530
4.18.3.2 Psychological Assessment as a Treatment Technique 530
4.18.3.3 Psychological Assessment for Outcomes Assessment 531

4.18.4 GENERAL CONSIDERATIONS FOR THE SELECTION AND USE OF PSYCHOLOGICAL TEST INSTRUMENTATION 531
4.18.4.1 Types of Instrumentation for Therapeutic Assessment 532
4.18.4.1.1 Psychological/psychiatric symptom measures 532
4.18.4.1.2 Measures of general health status and role functioning 533
4.18.4.1.3 Quality of life measures 534
4.18.4.1.4 Service satisfaction measures 534
4.18.4.2 Guidelines for Instrument Selection 535
4.18.4.2.1 National Institute of Mental Health criteria 535
4.18.4.2.2 Other criteria and considerations 536

4.18.5 PSYCHOLOGICAL ASSESSMENT AS A TOOL FOR SCREENING 538
4.18.5.1 Research-based Use of Psychological Screeners 539
4.18.5.2 Implementation of Screeners into the Daily Work Flow of Service Delivery 540

4.18.6 PSYCHOLOGICAL ASSESSMENT AS A TOOL FOR TREATMENT PLANNING 541
4.18.6.1 Assumptions About Treatment Planning 541
4.18.6.2 The Benefits of Psychological Assessment for Treatment Planning 542
4.18.6.2.1 Problem identification 542
4.18.6.2.2 Problem clarification 542
4.18.6.2.3 Identification of important patient characteristics 543
4.18.6.2.4 Monitoring of progress along the path of expected improvement 544

4.18.7 PSYCHOLOGICAL ASSESSMENT AS A THERAPEUTIC INTERVENTION 545
4.18.7.1 What Is Therapeutic Assessment? 545
4.18.7.2 The Impetus for Therapeutic Assessment 546
4.18.7.3 The Therapeutic Assessment Process 546
4.18.7.3.1 Step 1: The initial interview 547
4.18.7.3.2 Step 2: Preparing for the feedback session 547
4.18.7.3.3 Step 3: The feedback session 547
4.18.7.3.4 Additional steps 548
4.18.7.4 Empirical Support for Therapeutic Assessment 548

4.18.8 PSYCHOLOGICAL ASSESSMENT AS A TOOL FOR OUTCOMES MANAGEMENT 549
4.18.8.1 What Are Outcomes? 549
4.18.8.2 Outcomes Assessment: Measurement, Monitoring, and Management 550

4.18.8.3 The Benefits of Outcomes Assessment 550
4.18.8.4 The Therapeutic Use of Outcomes Assessment 550
4.18.8.4.1 Purpose of the outcomes assessment 551
4.18.8.4.2 What to measure 551
4.18.8.4.3 How to measure 552
4.18.8.4.4 When to measure 553
4.18.8.4.5 How to analyze outcomes data 554
4.18.9 FUTURE DIRECTIONS 555
4.18.9.1 What the Industry Is Moving Away From? 555
4.18.9.2 Trends in Instrumentation 556
4.18.9.3 Trends in Data Use and Storage 556
4.18.9.4 Trends in the Application of Technology 557
4.18.10 SUMMARY 558
4.18.11 REFERENCES 559

4.18.1 INTRODUCTION

The cost of health care in the USA has reached astronomical heights. In 1995, approximately $1 trillion, or 14.9% of the gross domestic product, was spent on health care, and a 20% increase is expected by the year 2000 (*Mental Health Weekly*, 1996a). The cost and prevalence of mental health problems and the accompanying need for behavioral health care services in the USA continue to rise at rates which give cause for concern. America's mental health bill in 1990 was $147 billion (*Mental Health Weekly*, 1996c). The Center for Disease Control and Prevention (1994) recently reported on the results of a survey of 45 000 randomly interviewed Americans regarding their quality of life. The survey found that one-third of the respondents reported they suffered from depression, stress, or emotional problems at least one day a month, and 11% percent of the sample reported having these problems more than eight days a month.

The American Psychological Association (APA; 1996) also reports statistics, summarized below, that bear attention.

(i) It is estimated that 15–18% of Americans suffer from a mental disorder; 14 million of these individuals are children.

(ii) Approximately eight million Americans suffer from depression in any given one-month period.

(iii) As many as 20% of Americans will suffer one or more major episodes of depression during their lifetime.

(iv) An estimated 80% of elderly residents in Medicaid facilities were found to have moderate to intensive needs for mental health services.

Moreover, information from various studies indicates that at least 25% of primary health care patients have a diagnosable behavioral disorder (*Mental Health Weekly*, 1996b).

The need for behavioral health care services is significant. In analyzing data from a 1987 national survey of 40 000 people in 16 000 households, Olfson and Pincus (1994a, 1994b) found that 3% of the population was seen for at least one psychotherapeutic session that year. Of these visits, 81% were to mental health professionals. Estimates provided by VandenBos, DeLeon, and Belar (1993) in the early 1990s indicated that in any year, 37.5 million Americans (or 15% of the population at that time) could benefit from mental health services.

What is the value of the services provided to those suffering from mental illness or substance abuse/addiction/dependency? Some might argue that the benefit is either minimal, or too costly to achieve if significant effects are to be gained. This is in the face of data which suggest otherwise. Numerous studies have demonstrated that treatment of mental health and substance abuse/dependency problems can result in substantial savings when viewed from a number of perspectives. This "cost offset" effect probably has been demonstrated most clearly in savings in medical care dollars over given periods of time.

Medical cost offset considerations are significant, given reports that 50–70% of usual primary care visits are for medical problems that involve psychological factors (APA, 1996). APA also reports that 25% of patients seen by primary care physicians have a disabling psychological disorder, and that depression and anxiety rank among the top six conditions seen by family physicians. Following are just a few of the findings supporting the medical cost benefits that can accrue from providing behavioral health care treatment.

(i) At least 25% or more of patients seen in a primary care setting have diagnosable behavioral disorders and use two to four times as many medical resources as those patients without these disorders (*Mental Health Weekly*, 1996b).

(ii) Sipkoff (1995) reported several conclusions, drawn from a review of numerous studies

conducted between 1988 and 1994 and listed in the *Cost of addictive and mental disorders and effectiveness of treatment* report published by the Substance Abuse and Mental Health Services Administration (SAMHSA). One conclusion derived from a meta-analysis of offset effect was that treatment for mental health problems results in an approximately 20% reduction in the overall cost of health care. The report also concluded that while alcoholics were found to spend twice as much on health care as those without abuse problems, one-half of the cost of substance abuse treatment is offset within one year by subsequent reductions in the combined medical cost savings for the patient and his or her family.

(iii) Strain et al. (1991) found that screening a group of 452 elderly hip fracture patients for psychiatric disorders prior to surgery and providing mental health treatment to the 60% of the sample needing treatment reduced total medical expenses by $270 000. The cost of the psychological/psychiatric services provided to this group was only $40 000.

(iv) Simmons, Avant, Demski, and Parisher (1988) compared the average medical costs for chronic back pain patients at a multidimensional pain center (providing psychological and other types of intervention) during the year prior to treatment to those costs of the year following treatment. The pretreatment costs per patient were $13 284 while post-treatment costs were $5596.

The reader is referred to Friedman, Sobel, Myers, Caudill, and Benson (1995) for a detailed discussion of various ways in which behavioral interventions can both maximize care to medical patients and achieve significant economic gains.

APA (1996) has very succinctly summarized what appears to be the prevalent findings of the medical cost offset literature.

> (i) Patients with mental disorders are heavy users of medical services, averaging twice as many visits to their primary care physicians as patients without mental disorders.
>
> (ii) When appropriate mental health services are made available, this heavy use of the system often decreases, resulting in overall health savings.
>
> (iii) Cost offset studies show a decrease in total health care costs following mental health interventions even when the cost of the intervention is included.
>
> (iv) In addition, cost offset increases over time, largely because ... patients continue to decrease their overall use of the health care system, and don't require additional mental health services. (p. 2)

Medical cost offset effects are relatively obvious and easy to measure. Benefits, financial and otherwise, that accrue from the treatment of mental health and substance abuse/dependency problems also can come in forms that may not be so obvious. One area in which treatment can have a tremendous impact is that of the workplace. For example, note a few of the facts assembled by APA (1996).

(i) In 1985 behavioral health problems resulted in over $77 billion in lost income to Americans.

(ii) California's stress-related disability claims totaled $350 million in 1989.

(iii) In 1980, alcoholism resulted in over 500 million lost work days in this country.

(iv) Major depression cost an estimated $23 billion in lost work days in 1990. In addition, individuals with this disorder are three times more likely than nondepressed individuals to miss time from work and four times more likely to take disability days.

(v) Of all subjects from 58 psychotherapy effectiveness studies focusing on the treatment of depression, 77% received significantly better work evaluations than depressed subjects who did not receive treatment.

(vi) Treatment resulted in a 150% increase in earned income for alcoholics and a 390% increase in income for drug abusers in one study of 742 substance abusers.

In related findings, anxiety disorders accounted for one-third of America's $147 billion mental health bill in 1990 (*Mental Health Weekly*, 1996c). And on another front, the former director of the Office of the National Drug Control Policy reported that for every dollar spent on drug treatment, America saves seven dollars in health care and criminal justice costs (*Substance Abuse Funding News*, 1995).

Society's need for behavioral health care services provides an opportunity for trained providers of mental health services to become part of the solution to a major health care problem that shows no indication of decline. Each of the helping professions has the potential to make a particular contribution to this solution. Not the least of these contributions are those that can be made by clinical psychologists. As pointed out in an earlier volume (Maruish, 1994), the use of psychological tests in the assessment of the human condition is one of the hallmarks of clinical psychology. In fact, the training and acquired level of expertise in psychological testing distinguishes the clinical psychologist from other behavioral health care professionals probably more than anything else. Indeed, expertise in test-based psychological assessment can be said to be the particular and unique contribution that clinical psychologists make to the behavioral health care field.

For decades, clinical psychologists and other behavioral health care providers have come to rely on psychological assessment as a standard tool to be used with other sources of information for diagnostic and treatment planning purposes. However, changes that have taken place in the delivery of health care in general, and behavioral health care services in particular, during the past several years have led to changes in the way in which third-party payers and clinical psychologists themselves think about and/or use psychological assessment in day-to-day clinical practice. Some question the value of psychological assessment in the current time-limited, capitated service delivery arena where the focus has changed from clinical priorities to fiscal priorities (Sederer, Dickey, & Hermann, 1996). Others argue that it is in just such an arena that the benefits of psychological assessment can be most fully realized and contribute significantly to the delivery of cost-effective treatment for behavioral health disorders. Consequently, it could assist the health care industry in appropriately controlling or possibly reducing the utilization and cost of health care over the long term. It is this latter side of the argument that is supported by this author, and it provides the basis for this chapter.

In developing this chapter, the intent has been to provide students and practitioners of clinical psychology with an overview of how psychological assessment could and should be used in this era of managed behavioral health care. In doing so, this author discusses how psychological assessment is currently being used in the therapeutic environment and the many ways in which it might be used to the ultimate benefit of patients, providers, and payers.

As a final introductory note, it is important for the reader to understand that the term "psychological assessment," as it is used in this chapter, refers to the evaluation of a patient's mental health status using psychological tests or related instrumentation. This evaluation may be conducted with or without the benefit of patient or collateral interviews, review of medical or other records, and/or other sources of relevant information about the patient.

4.18.2 THE CURRENT PRACTICE OF PSYCHOLOGICAL ASSESSMENT IN THE THERAPEUTIC ENVIRONMENT

For a number of decades, psychological assessment has been viewed as a valued and integral part of the services offered by clinical psychologists. However, its popularity has not been without its ups and downs. Megargee and Spielberger (1992) have described a decrease in interest in assessment that began in the 1960s. This was due to a number of factors, including shifts in focus to those aspects of treatment for which assessment was thought to contribute little. Examples of these aspects included a growing emphasis on behavior modification techniques, the increasing use of psychotropic medications, and an emphasis in studying symptoms rather than personality syndromes and structures. Fortunately, Megargee and Spielberger also noted a number of factors that indicate a relatively recent resurgence in the interest in assessment, including a new realization of how psychological assessment can assist in interventions provided to mental health care patients.

But where does psychological assessment actually fit into the daily scope of activities for practicing psychologists? The results of two recent surveys provide inconsistent findings. The newsletter *Psychotherapy Finances* (1995) reported the results of a nationwide readership survey of 1700 mental health providers of various professions. In this survey, 67% of the participating psychologists reported that they provide psychological testing services. This represents about a 10% drop from a similar survey published in 1992 by the same publication. Also of interest in this survey is the percent of professional counselors (39%), marriage and family counselors (16%), psychiatrists (21%), and social workers (13%) offering these same services.

In a 1995 survey conducted by the APA's Committee for the Advancement of Professional Practice (Phelps, 1996), 14000 practitioners responded to questions related to workplace settings, areas of practice concerns, and range of activities. Most of the respondents (40.7%) were practitioners whose primary work setting was an individual independent practice. Other general work settings, that is, government, medical, academic, group practice settings, were represented by fairly equal numbers of respondents from the remainder of the sample. The principal professional activity reported by the respondents was psychotherapy, with 43.9% of the sample acknowledging involvement in this service. Assessment was the second most prevalent activity, being reported by 14% of the sample.

Differences in the two samples utilized in the above surveys may account for the inconsistencies in their findings. Psychologists who are subscribers to *Psychotherapy Finances* may represent that subsample of the APA survey respondents who are more involved in the delivery of clinical services. Certainly the fact that only about 44% of the APA respondents offer psychotherapy services supports this hypothesis.

Regardless of the two sets of findings, psychological assessment does not appear to be utilized as much as in the past, and one does not have to look hard to determine at least one reason why. One of the major changes that has come about in the American health care system during the past several years has been the creation and proliferation of managed care organizations (MCOs). The most significant direct effects of managed care include reductions in the length and amount of service, reductions in accessibility to particular modalities (e.g., reduced number of outpatient visits per case), and profession-related changes in the types of services managed by behavioral health care providers (Oss, 1996). Overall, the impact of managed behavioral health care on the services offered by psychologists and other health care providers has been tremendous. In the APA survey reported above (Phelps, 1996), approximately 79% of the respondents reported that managed care had either a low, medium, or high negative impact on their work. How has managed care negatively impacted the use of psychological assessment? It is not clear from the results of this survey, but perhaps others can offer at least a partial explanation.

Ficken (1995) has provided some insight into how the advent of managed care has limited the reimbursement for (and therefore the use of) psychological assessment. In general, he sees the primary reason for this as being a financial one. In an era of capitated behavioral health care coverage, the amount of money available for behavioral health care treatment is limited. MCOs therefore require a demonstration that the amount of money spent for testing will result in a greater amount of treatment cost savings. In addition, Ficken notes that much of the information obtained from psychological assessment is not relevant to the treatment of patients within an MCO environment. Understandably, MCOs are reluctant to pay for the gathering of such information.

Werthman (1995) provides similar insights into this issue, noting that

> Managed care ... has caused [psychologists] to revisit the medical necessity and efficacy of their testing practices. Currently, the emphasis is on the use of highly targeted and focused psychological and neuropsychological testing to sharply define the "problems" to be treated, the degree of impairment, the level of care to be provided and the treatment plan to be implemented.
>
> The high specificity and "problem-solving" approach of such testing reflects MCOs' commitment to effecting therapeutic change, as opposed to obtaining a descriptive narrative with scores. In this context, testing is perceived as a strong tool for assisting the primary provider in more accurately determining patient "impairments" and how to "repair" them. (p. 15)

In general, Werthman views psychological assessment as being no different from other forms of patient care, thus making it subject to the same scrutiny, demands for demonstrating medical necessity and/or utility, and consequent limitations imposed by MCOs on other covered services.

The foregoing representations of the current state of psychological assessment in behavioral health care delivery could be viewed as an omen of worse things to come. In this author's opinion, they are not. Rather, the limitations that are being imposed on psychological assessment and the demand for justification of its use in clinical practice represent part of the customers' dissatisfaction with the way things always have been done in the past. In general, this author views the tightening of the purse strings as a positive move for both behavioral health care and the profession of psychology. It is a wake-up call to those who have contributed to the health care crisis by either uncritically performing costly psychological assessments, being unaccountable to the payers and recipients of our services, and generally not performing our services in the most responsible, cost-effective and efficient way possible. It is telling us that we need to evaluate what we've done and the way we've done it, and to determine what is the best way to do it in the future. As such, it provides an opportunity for clinical psychologists to re-establish the valuable contributions they can make to improving the quality of behavioral health care delivery through their knowledge and skills in the area of psychological assessment.

In the sections that follow, this author will convey what he sees are the opportunities for psychological assessment in the behavioral health care arena, both in the present and the future, and the means of best achieving them. The views that are advanced are based on his knowledge of and experience in current psychological assessment practices as well as directions provided by the current literature. Some will probably disagree with the proposed approach, given their own experience and thinking on the matters discussed. However, it is hoped that even though in disagreement, the reader will be challenged to defend his or her position to themselves and as a result, feel more comfortable in their thinking about their approach to their psychological assessment practices.

4.18.3 PSYCHOLOGICAL ASSESSMENT AS A THERAPEUTIC ADJUNCT

The role of psychological assessment in the therapeutic environment traditionally has been quite limited. Those of us who did not receive our graduate clinical training within the past few years probably have been taught the value of psychological assessment only at the "front end" of treatment. We were instructed in the power and utility of psychological assessment as a means of assisting in the identification of symptoms and their severity, personality characteristics relevant to understanding the patient and his or her typical way of perceiving and interacting with the world, and other aspects of the individual (e.g., intelligence, vocational interests) that are important in arriving at a description of the patient at one particular point in time. Based on these data and information obtained from patient and collateral interviews, medical records and the individual's stated goals for treatment, a diagnostic impression was given and a treatment plan was probably formulated and placed in the patient's chart, hopefully to be reviewed at various points during the course of treatment. In some cases, the patient was assigned to another practitioner within the same organization or referred out, never to be contacted or seen again, much less be assessed again by the one who performed the original assessment.

Fortunately, during the past few years the usefulness of psychological assessment as more than just a tool to be used at the beginning of treatment has come to be recognized. Consequently, its utility has been extended beyond being a mere tool for describing an individual presenting themselves for treatment to being a means of facilitating the treatment and understanding of behavioral health care problems throughout the episode of care and beyond. Psychologists and others who employ it in their practices are now finding that psychological assessment can be used for a variety of purposes. Generally speaking, several psychological tests currently being marketed can be employed as tools for assisting in clinical decision-making, outcomes assessment and, more directly, as treatment techniques in and of themselves. Each of these uses can uniquely contribute incremental value to the therapeutic process.

4.18.3.1 Psychological Assessment for Clinical Decision-making

Traditionally, psychological assessment has been used to assist clinical psychologists and other behavioral health care clinicians in making important clinical decisions. The types of decision-making for which it has been used include those related to screening, treatment planning, and monitoring of treatment progress. Generally, screening may be undertaken to assist in either: (i) identifying the patient's need for a particular service, or (ii) determining the likelihood of the presence of a particular disorder or other behavioral/emotional/psychological problem. More often than not, a positive finding on screening leads to a more extensive evaluation of the patient in order to confirm with greater certainty the existence of the problem, or to further delineate the problem. The value of screening lies in the fact that it permits the clinicians to identify, quickly and economically, with a fairly high degree of confidence (depending on the particular instrumentation used), those who are and are not likely to need care or at least further evaluation.

In many instances, psychological assessment is performed in order to obtain information that is deemed useful in the development of a specific plan for treatment. Typically, it is the type of information that is not easily (if at all) accessible through other means or sources. It is information which, when combined with other information about the patient, aids in understanding the patient, identifying the most important problems and issues that need to be addressed, and formulating recommendations about the best means of addressing them.

Another way in which psychological assessment can play a role in clinical decision-making is in the area of treatment monitoring. Repeated assessment of the patient at regular intervals during the treatment process can provide the therapist with feedback regarding the progress which is being made in the therapeutic endeavor. Based on the findings, the therapist will be encouraged either to continue with the original therapeutic approach or, in the case of no change or exacerbation of the problem, to modify or abandon the approach in favor of an alternate one.

4.18.3.2 Psychological Assessment as a Treatment Technique

It is only recently that empirical studies and other articles addressing the therapeutic benefits that can be realized directly from discussing psychological assessment results with the patient have been published. Rather than just providing test feedback as directed by APA's *Ethical principles of psychologists* (APA, 1992), therapeutic use of assessment involves a presentation of assessment results (including assessment materials such as test protocols, profile forms, other assessment summary materials) directly to the patient; an elicitation of the patient's reactions to them; and an in-depth discussion of the meaning of the results in terms of patient-defined assessment goals. In essence, the assessment data can serve as a catalyst for the therapeutic encounter via the objective feedback that is provided to the patient, the patient self-assessment that is stimulated, and the opportunity for patient and therapist to arrive at mutually agreed upon therapeutic goals, based on impressionistic and objective data available to both parties.

4.18.3.3 Psychological Assessment for Outcomes Assessment

Currently, one of the most common reasons for conducting psychological assessment in the USA is to assess the outcomes of behavioral health care treatment. It is difficult to open a trade paper or health care newsletter or to attend a professional conference without being presented with a discussion on either how to "do outcomes" or what the results of a certain facility's outcomes study have revealed. The focus on outcomes assessment most probably can be traced to the "continuous quality improvement" (CQI) movement that was initially implemented in business and industrial settings. The impetus for the movement originally was a desire to produce quality products in the most efficient manner, resulting in increased revenues and decreased costs.

In the health care arena, outcomes assessment has multiple purposes, not the least of which is as a tool for marketing the organization's services. Related to this, those organizations vying for lucrative contracts from third-party payers to provide health care services to their covered lives frequently require outcomes data demonstrating the effectiveness of the services offered by the bidders. Equally important to those awarding contracts is how satisfied patients are with the provider's services. But probably the most important potential use of this data for provider organizations (although not always recognized as such) can be found in the knowledge it yields about what works and what doesn't. In this regard it can serve a program evaluation function. It is this knowledge that, if attended to and acted upon, can lead to improvement in the services the organization offers. When used in this manner, outcomes assessment can become an integral component of the organization's CQI initiative.

But more importantly for the individual patient, outcomes assessment provides a means of objectively measuring how much improvement he or she has made from the time of treatment initiation to the time of treatment termination. Feedback to this effect may serve to instill in the patient greater self-confidence and self-esteem, and/or a more realistic view of where he or she is (from a psychological standpoint) at that particular time in their life. Conversely, it may serve as an objective indicator to the patient of the need for continued treatment.

The purpose of the foregoing is to present a broad overview of psychological assessment as a multipurpose behavioral health care tool. Depending on the individual clinician or provider organization, it may be employed for one or more, or all, of the purposes just described. Knowing the various ways in which psychological assessment can be used in the service of therapeutic change should help the reader understand the more in-depth and detailed discussion about how these applications can facilitate or otherwise add value to the psychotherapeutic services offered by providers. This detailed discussion follows below.

Before beginning this discussion, however, it is important to briefly review the types of instrumentation most likely to be used in therapeutic psychological assessment, as well as the significant considerations and issues related to the selection and use of this instrumentation for the stated purposes. This should further facilitate the reader's understanding of the remainder of the chapter.

4.18.4 GENERAL CONSIDERATIONS FOR THE SELECTION AND USE OF PSYCHOLOGICAL TEST INSTRUMENTATION

New instrumentation for facilitating and evaluating behavioral health care treatment is released by the major test publishers annually. Thus, the availability of instrumentation for these purposes is not an issue. However, selection of the appropriate instrument(s) for one or more of the therapeutic purposes described above is a matter requiring careful consideration. Inattention to the instrument's intended use, its demonstrated psychometric characteristics, its limitations, and other aspects related to its practical application can result in misguided treatment and potentially harmful consequences for a patient.

Several types of instruments could be used for the general therapeutic assessment purposes described above. For example, neuropsychological instruments might be used to assess memory deficits that could impact the clinician's decision to perform further testing, the goals established for treatment, and the approach to treatment that is selected. Tests designed to provide estimates of level of

intelligence might be used for the same purposes. It is beyond the scope of this chapter to address, even in the most general way, all of the types of tests, rating scales, and the instrumentation that might be employed in the therapeutic environment. Instead, the focus here will be on general classes of instrumentation that have the greatest applicability in the service of the therapeutic endeavor. To a limited extent, specific examples of such instruments will be presented. This will be followed by a discussion of criteria and considerations that will assist the clinician in selecting the best instrumentation for his or her intended purposes.

4.18.4.1 Types of Instrumentation for Therapeutic Assessment

The instrumentation required for any therapeutic application will depend on: (i) the general purpose(s) for which the assessment is being conducted, and (ii) the level of informational detail that is required for those purpose(s). Generally, one may classify the types of instrumentation that would serve the purpose(s) of the therapeutic assessment into one of four general categories. As mentioned above, other types of instrumentation are frequently used in clinical settings for therapeutic purposes. However, the present discussion will be limited to those more commonly used by a wide variety of clinical psychologists in their day-to-day practices.

4.18.4.1.1 Psychological/psychiatric symptom measures

Probably the most frequently used instrumentation for several therapeutic purposes are measures of psychopathological symptomatology. Besides the fact that these are the types of instruments on which the majority of the clinician's psychological assessment training has probably been focused, they were developed to assess the problems that typically prompt people to seek treatment.

There are several subtypes of these measures of psychological/psychiatric symptomatology. The first is the comprehensive multidimensional measure. This is typically a lengthy, multiscale instrument that measures and provides a graphical profile of the patient on several types of psychopathological symptom domains (e.g., anxiety, depression) or disorders (schizophrenia, antisocial personality). Also, summary indices sometimes are available to provide a more global picture of the individual with regard to his or her psychological status or level of distress. Probably the most widely used and/or recognized of these measures are the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1951) and its restandardized revision, the MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989), the Millon Clinical Multiaxial Inventory-III (MCMI-III; Millon, 1994), and the Personality Assessment Inventory (PAI; Morey, 1991).

Multiscale instruments of this type can serve a variety of purposes that facilitate therapeutic efforts. They may be used upon initial contact with the patient to screen for the need for service and, at the same time, yield information that is useful for treatment planning. Indeed, some such instruments (e.g., the MMPI-2) may make available supplementary, content-related, and/or special scales that are designed to assist the user in addressing specific treatment considerations (e.g., low motivation for treatment). Other multiscale instruments might be useful in identifying specific problems that may be unrelated to the patient's chief complaints (e.g., low self-esteem). They can also be administered at numerous times during the course of treatment to monitor the patient's progress toward achieving established goals and to assist in determining what adjustments (if any) must be made to the clinician's approach. In addition, use of the instrument in a pre- and post-treatment fashion provides information related to the outcomes of the treatment. Data obtained in this fashion can be analyzed with results from other patients to evaluate the effectiveness of an individual therapist as well as an organization.

Abbreviated multidimensional measures are quite similar to the comprehensive multidimensional measure in many respects. First, by definition, they contain multiple scales for measuring a variety of symptom domains and/or disorders. They also may allow for the derivation of an index of the patient's general level of psychopathology or distress. In addition, they may be used for screening, treatment planning and monitoring, and outcomes assessment purposes just like the comprehensive instruments. The distinguishing feature of the abbreviated instrument is its length. Again, by definition, these instruments are relatively short, and easy to administer and (usually) score. Their brevity does not allow for an in-depth assessment of the patient and his or her problems, but this is not what these instruments were designed to do.

Probably the most widely used of these brief instruments are Derogatis' family of symptom checklist instruments. These include the original Symptom Checklist-90 (SCL-90; Derogatis, Lipman, & Covi, 1973) and its revision, the SCL-90-R (Derogatis, 1983). Both of these instruments contain a checklist of 90 psychological symptoms, most of which score on the instruments' nine symptom scales. For each of these instruments an even briefer version has been developed. The first is the Brief Symptom Inventory (BSI; Derogatis, 1992), which was derived from the SCL-90-R. In a health care environment that is cost-conscious and unwilling to make too many demands on patient time, this 53-item instrument is gaining popularity over its longer and more expensive 90-item parent instrument. Similarly, a brief form of the original SCL-90 has been developed. Titled the Symptom Assessment-45 Questionnaire (SA-45; Strategic Advantage, Inc., 1996), its development did not follow Derogatis' approach to the development of the BSI; instead, cluster analytic techniques were used to select five items each for assessing each of the nine symptom domains found on the three Derogatis checklists.

The major strength of the abbreviated multiscale instruments is their ability to broadly and very quickly survey several psychological symptom domains and/or disorders relative to the patient. Its value is most clearly evident in settings where both the time and dollars available for assessment services are quite limited. These instruments provide a lot of information quickly. Because of their brevity, they are much more likely to be completed by patients than their lengthier comprehensive counterparts. This last point is particularly important if one is interested in monitoring treatment or assessing outcomes, both of which require at least two or more assessments to obtain the desired information.

4.18.4.1.2 Measures of general health status and role functioning

During the past decade, there has been an increasing interest in the assessment of health status in health care delivery systems. Initially, this interest was shown mostly by those organizations and settings focusing primarily on the treatment of physical diseases and disorders. Within recent years, behavioral health care providers have recognized the value in assessing the patient's general level of health. It is important to recognize that the term "health" means more than just the absence of disease or debility; it also implies a state of well-being throughout the individual's physical, psychological, and social spheres of existence (World Health Organization [WHO], 1948). Dickey and Wagenaar (1996) point out how this view of health recognizes the importance of eliciting the patient's point of view in assessing health status. They also point to similar conclusions reached by Jahoda (1958) specific to the area of mental health. Here, an individual's self-assessment relative to how he or she feels they should be is an important component of "mental health."

Measures of health status and physical functioning can be classified into one of two groups: generic and condition-specific. Probably the most widely used and respected generic health status measures are the 36-item Medical Outcomes Study Short Form Health Scale (SF-36; Ware & Sherbourne, 1992; Ware, Snow, Kosinski, & Gandek, 1994) and the 39-item Health Status Questionnaire 2.0 (HSQ; Health Outcomes Institute, 1993; Radosevich, Wetzler, & Wilson, 1994). Aside from the minor variations in the scoring of one of the instruments' scales (i.e., Bodily Pain) and the HSQ's inclusion of three depression screening items, the two measures essentially are identical. Each assesses eight dimensions of health, four addressing mental health-related constructs and four addressing physical health-related constructs, that reflect the WHO concept of "health."

Role functioning has recently gained attention as an important variable to address in the course of assessing the impact of a physical or mental disorder on an individual. In devising a treatment plan and monitoring progress over time, it is important to know how the person's ability to work, perform daily tasks, or interact with others is affected by the disorder. The SF-36 and HSQ both address these issues with scales designed for this purpose.

Responding to concerns that even these relatively brief objective measures are too lengthy for regular administration in clinical and research settings, 12-item, abbreviated versions of each have been developed. The SF-12 (Ware, Kosinski, & Keller, 1995) was developed for use in large scale, population-based research where the monitoring of health status at a broad level is all that is required. Also, a 12-item version of the HSQ, the HSQ-12 (Radosevich & Pruitt, 1996), was developed for similar uses. Interestingly, given that the two abbreviated versions were derived from essentially the same instrument, there is only a 50% item overlap between the two shortened instruments. Both instruments are relatively new but the data supporting their use that has been gathered up to 1997 is promising.

Condition-specific health status and functioning measures have been utilized for a number of years. Most have been developed for use with physical rather than mental disorders, diseases, and conditions. However, condition-specific measures of mental health

status and functioning are beginning to appear. A major source of this type of instrument is the Minnesota-based Health Outcomes Institute (HOI), a successor to the health care think tank InterStudy. In addition to the HSQ and the HSQ-12, HOI serves as the distributor/clearinghouse for the condition-specific "technology of patient experience (TyPE) specifications." The available TyPEs that would be most useful to clinical psychologists and other behavioral health care practitioners include those developed by a team of researchers at the University of Arkansas Medical Center for use with depressive, phobic, and alcohol and substance disorders. TyPEs for other specific psychological disorders are currently under development at the University of Arkansas for distribution through HOI.

4.18.4.1.3 Quality of life measures

In their brief summary of this area, Andrews, Peters, and Teesson (1994) indicate that most of the definitions of "quality of life" (QOL) describe a multidimensional construct encompassing physical, affective, cognitive, social, and economic domains. Objective measures of QOL focus on environmental resources required to meet one's needs and can be completed by someone other than the patient. The subjective measures of QOL assess the patient's satisfaction with the various aspects of his or her life and thus must be completed by the patient.

Andrews et al. (1994) draw other distinctions in the QOL arena. One has to do with the differences between QOL and health-related quality of life, or HRQL, and (similar to the case with health status measures) the other has to do with the distinction between generic and condition-specific measures of QOL. QOL measures differ from HRQL measures in that the former assess the whole "fabric of life," while the latter assess quality of life as it is affected by a disease or disorder, or by its treatment. Generic measures are designed to assess aspects of life that are generally relevant to most people; condition-specific measures are focused on aspects of the lives of particular disease/disorder populations. However, as Andrews et al. point out, generic and condition-specific QOL measures tend to overlap quite a bit.

4.18.4.1.4 Service satisfaction measures

With the exploding interest in assessing the outcomes of treatment for the patient, it is not surprising to see an accompanying interest in assessing the patient's and, in some instances, the patient's family's satisfaction with the services received. In fact, many professionals and organizations equate satisfaction with outcomes and frequently consider it the most important outcome. In a recent survey of 73 behavioral health care organizations, 71% of the respondents indicated that their outcomes studies included measures of patient satisfaction (Pallak, 1994).

Although some view service satisfaction as an outcome, it is this author's contention that it should not be classified as such. Rather, it should be considered a measure of the overall therapeutic process, encompassing the patient's (and at times, others') view of how the service was delivered, the capabilities and attentiveness of the service provider, the benefits of the service (if any), and any of a number of other selected aspects of the service he or she received. Patient satisfaction surveys don't answer the question "What was the result of the treatment rendered to the patient"; they do answer the question "How did the patient feel about the treatment he or she received?" Thus, they serve an important program evaluation/improvement function.

The number of questionnaires that are currently being used to measure patient satisfaction is countless. This reflects the attempts of individual health care organizations to develop customized measures that assess variables important to their particular needs, which in turn reflects a response to outside demands to "do something" to demonstrate the effectiveness of their services. Often, this "something" has not been evaluated to determine its basic psychometric properties. As a result, there exists numerous options that one may choose from, but very few that actually have demonstrated their validity and reliability as measures of service satisfaction.

Fortunately, there are a few instruments that have been investigated for their psychometric integrity. Probably the most widely used and researched patient satisfaction instrument designed for use in behavioral health care settings is the eight-item version of the Client Satisfaction Questionnaire (CSQ-8; Attkisson & Zwick, 1982; Nguyen, Attkisson, & Stenger, 1983). The CSQ-8 was derived from the original 31-item CSQ (Larsen, Attkisson, Hargreaves, & Nguyen, 1979), which also yielded two longer 18-item alternate forms, the CSQ-18A and CSQ-18B (LeVois, Nguyen, & Attkisson, 1981). The more recent work of Attkisson and his colleagues at the University of California at San Francisco is the Service Satisfaction Scale-30 (SSS-30; Greenfield & Attkisson, 1989), a 30-item multifactorial scale that yields information regarding different aspects of satisfaction with mental health service, such as perceived outcome and manner and skill of the clinician.

4.18.4.2 Guidelines for Instrument Selection

Regardless of the type of instrument one might consider using in the therapeutic environment, many clinical psychologists frequently must choose between many product offerings. But what are the general criteria for the selection of any instrument for psychological assessment? What should guide the clinician's selection of an instrument for a specific therapeutic purpose? As part of their training, clinical psychologists and professionals from related psychological specialties have been educated about the important psychometric properties that should be considered when determining the appropriateness of an instrument for its intended use. However, this is just one of several issues that should be taken into account in an evaluation of a specific instrument for a specific therapeutic use. The guidance that has been offered by experts with regard to instrument selection is worth noting here.

4.18.4.2.1 National Institute of Mental Health criteria

Probably the most thorough and clinically relevant guidelines for the selection of psychological assessment instruments comes from the National Institute of Mental Health (NIMH) supported work of Ciarlo, Brown, Edwards, Kiresuk, and Newman (1986). A synopsis of Newman and Ciarlo's (1994) updated summary of this NIMH work is presented here. Note that the criteria discussed below were originally developed for use in evaluating instruments for outcomes assessment purposes. However, most have relevance to the selection of instrumentation used for the other therapeutic assessment purposes described above. Exceptions and qualifications with regard to this issue will be noted when appropriate.

Newman and Ciarlo (1994) describe 11 criteria for the selection of outcomes assessment instruments, each of which can be grouped into one of five types of consideration. The first consideration is that of applicability. The issue here is the relevance of the instrument to the target population. The instrument should assess those problems, symptoms, characteristics, and so on, that are common to the group to whom the instrument will be administered. The more heterogeneous the population, the more chance that modifications will be required and that these will alter the standardization and psychometric integrity of the instrument. Another applicability issue to consider when the instrument is to be used for outcomes assessment purposes is its independence from the type of treatment to be offered to the population.

The second set of general considerations is that of methods and procedures (Newman & Ciarlo, 1994). Several selection criteria are related to this group. The first is that administration of the instrument is simple and easily taught. Generally, this is more of an issue with clinician-rating scales than self-report scales. In the case of rating scales, concrete examples, or objective referents, at each rating level should be provided to the user. Next, the instrument should allow input not only from the patient but also from other sources (e.g., the clinician, collaterals). The benefits of this include the opportunities to obtain a feel for the patient from many perspectives, to validate reported findings and observations, and to promote honesty in responding from all sources (given that all parties will know that others will also be providing input). The final methods and procedures criterion, though not necessarily as important for the instrument being used for screening or treatment planning purposes, is that the instrument provide information relevant to understanding how the treatment may have effected change in the individual.

Newman and Ciarlo's (1994) third set of considerations have to do with the psychometric strengths of the instruments. According to the NIMH panel of experts, outcomes measures should: (i) meet the minimum psychometric standards for reliability (including internal consistency, test–retest reliability, and as appropriate, interrater reliability) and validity (content, construct, and concurrent validity); (ii) be difficult to "fake bad" or "fake good"; and (iii) be free from response bias and not reactive or sensitive to factors unrelated to the constructs that are being measured (e.g., physical settings, behavior of the treatment staff). These criteria obviously also apply to other psychological instruments used for purposes other than outcomes assessment. However, for outcomes assessment purposes, the instrument also must be sensitive to change related to treatment.

The fourth group of considerations concerns the cost of the instruments. Newman and Ciarlo (1994) point out that the answer to the question of how much one should spend on assessment instrumentation and associated costs (e.g., staff time for administering, scoring, processing, and analyzing the data) will depend on how important the data gathered is to assuring a positive return on the functions they support. In the context of the NIMH undertaking, Newman and Ciarlo felt that the data obtained through treatment outcomes assessment would support screening/treatment planning, efforts in quality assurance and program

evaluation, cost containment/utilization review activities, and revenue generation efforts. However, that may be considered the ideal. At this point, the number and nature of the purposes that would be supported by the obtained data will depend on the individual organization. The more purposes the data can serve, the less costly the instrumentation is likely to be, at least from a value standpoint. In terms of actual costs, Ciarlo et al. (1986) estimated that 0.5% of an organization's total budget would be an affordable amount for materials, staff training, data collection, and processing costs related to outcomes assessment. However, one should be mindful that the recommendation was made in 1986 and may not reflect changes in policies, requirements, and attitudes related to the use of psychological assessment instruments since that time.

The final set of considerations in instrument selection has to do with the utility of the instrument. Four criteria related to utility are posited by Newman and Ciarlo (1994). First, the scoring procedures and the manner in which the results are presented should be comprehensible to all with a stake in the treatment of the organization's patients. This would not only include the patient, his or her family, the organization's administrative staff and other treatment staff, but also third-party payers and (in the case of outcomes assessment or program evaluation) legislative and administrative policy makers. Related to this is the criterion that the results of the instrument be easily interpreted by those with a stake in them. Another utility-related criterion is that the instrument should be compatible with a number of clinical practices and theories that are employed in the behavioral health care arena. This should allow for a greater range of test applicability and greater acceptance by the various stakeholders in the patient's treatment.

Another important aspect of utility is that "the instrument support[s] the clinical processes of a service with minimal interference" (Newman & Ciarlo, 1994, p. 107). There are two issues here. The first has to do with whether the instrument can support the screening, planning, and/or monitoring activities in addition to the outcomes assessment activities. In other words, are multiple purposes served by the instrument's results? The second issue is one that has to do with the extent to which the organization's staff is burdened with the collection and processing of assessment data. How much will the assessment process interfere with the daily work flow of the organization's staff? Equally important is whether the benefits that accrue justify the cost of implementing an assessment program for whatever purpose(s).

4.18.4.2.2 Other criteria and considerations

Although the work of Ciarlo and his colleagues provides more extensive instrument selection guidelines than most, others who have addressed the issue have arrived at recommendations that serve to reinforce and/or complement those found in the NIMH document. For example, Gavin Andrews' work in Australia has led to significant contributions to the body of outcomes assessment knowledge. As part of this, Andrews et al. (1994) have identified six general "qualities of consumer outcome measures" that are generally in concordance with those from the NIMH study. First, the measure should meet the criterion of applicability. In other words,

> it should address dimensions which are important to the consumer (symptoms, disability, and consumer satisfaction) and useful for the clinician in formulating and conducting treatment, yet the measure should be one which can have its data aggregated in a meaningful way so that the requirements of management can be addressed. (p. 30)

Multidimensional instruments yielding a profile of scores on all dimensions of interest are viewed as a means of best serving the interests of all concerned.

Acceptability, that is, being both brief and user-friendly, is another desirable quality identified by Andrews et al (1994). Closely associated with this is the criterion of practicality. It might be viewed as a composite of those NIMH criteria related to matters of cost, ease of scoring and interpretation, and training in the use and interpretation of the measure. Again in agreement with the NIMH work, the final three criteria identified by Andrews et al. relate to reliability, validity, and sensitivity to change. With regard to reliability, Andrews et al. specify what they consider to be the minimum levels of acceptable internal consistency reliability (0.90 for long tests), interrater reliability (0.40), and construct and criterion validity (0.50). They also stress the importance of an instrument's face validity in helping to ensure cooperation from the patient, and of self-report instruments having multiple response options (rather then just "yes/no" options) for increasing sensitivity of an instrument to small but relevant changes in the patient's status over time.

In Ficken's (1995) discussion of the role of assessment in an MCO environment, he concludes that the difficulties clinicians are experiencing in demonstrating the utility of psychological assessment to payers lies in the fact the instruments and objectives of traditional psychological assessment are not in synch

with the needs of MCOs. The solution to the problem appears simple:

> the underlying objectives of testing must be aligned with the values and processes of MCOs. In short, this means identifying decision points in managed care processes that could be improved with objective, standardized data. There are two avenues in which these can be pursued: through facilitation/objectification of clinical-decision processes and through outcome assessment. (p. 12)

In general, Ficken (1995) sees opportunities in areas that this author has previously identified as screening, treatment planning, and outcomes assessment, specifically in the areas of primary medical care and behavioral health care (see below). Requirements of instruments used for screening were noted to include:

(i) high levels of sensitivity and specificity to diagnostic criteria from the *Diagnostic and statistical manual of mental disorders* (4th ed., *DSM-IV*; American Psychiatric Association, 1994) or the most up-to-date version of the *International classification of diseases* (*ICD*);

(ii) a focus on hard-to-detect (in a single office visit) but treatable disorders that are associated with imminent harm to self or others, significant suffering, and a decrease in productivity;

(iii) an administration time of no more than 10 minutes; and

(iv) an administration protocol that easily integrates into the organization's work flow.

Cases testing "positive" on the screener would be administered one or more "second-tier" instrument(s) to establish severity and a specific diagnosis. Ficken feels that if they are to be accepted by MCOs, these second-tier instruments should meet the requirements of screeners and either specify or rule out a diagnosis.

According to Ficken (1995), successful outcomes assessment instruments also must possess certain qualities. Because the areas most important to assess for outcomes measurement purposes are symptom reduction, level of functioning, quality of life and patient satisfaction, the instrument should (i) focus on one of these areas, (ii) be brief, (iii) meet "traditional standards" for validity and reliability, and (iv) be sensitive to clinical change.

Based on the work of Vermillion and Pfeiffer (1993), Burlingame, Lambert, Reisinger, Neff, and Mosier (1995) recommended four criteria for the selection of outcomes measures. The first is acceptable "technical features," that is, validity and reliability. Specifically, these authors recommended that instruments have an internal consistency of at least 0.80, test–retest reliability of at least 0.70, and concurrent validity of at least 0.50. The second criterion is "practicality features." These include brevity, ease of administration and scoring, and simplicity of the reporting of results. Third, the instrumentation should be "suitable" for the patients that are seen within the setting. Thus, because of the nature of most presenting problems in mental health settings, it should assess symptomatology and psychosocial functioning. The fourth criterion is sensitivity to "meaningful" change over time, allowing for a differentiation of *symptomatic* change from interpersonal/social role *functional* change.

Schlosser (1995) proposed a rather nontraditional view of "outcomes assessment." In what he refers to as a "patient-centric" view, assessment information is gathered and used during the course of therapy to bring about change during therapy, not after therapy has ended. Essentially, this equates to what this author has referred to above (and discusses in more detail below) as treatment monitoring. In this model, Schlosser feels that this type of assessment requires "elements regarding very specific, theoretically derived, empirically validated areas of functioning" (Schlosser, 1995, p. 66). These would involve the use of both illness and well-being measures that assess the patient on emotional, mental/cognitive, physical, social, life direction, and life satisfaction dimensions.

Many of Schlosser's (1995) considerations for selection of such measures are not unique (i.e., having "acceptable' levels of reliability and validity, brief, low-cost, and sensitive). However, for the purposes described Schlosser also indicates that they should also: (i) have "paradigmatic sensibility" (i.e., key words have the same meaning across instruments); (ii) be designed for repeated administration for feedback or self-monitoring purposes; and (iii) provide actionable information.

In addition to some already mentioned criteria (acceptable validity, reliability, affordability, ease of administration, and ease of data entry and analysis), Sederer et al. (1996) discuss other considerations that warrant attention in selecting outcomes measures for specific situations. These include automation capabilities related to availability of software for data analysis and reporting, compatibility with the organization's existing information system, and the ability to enter data via an optical scanner. They also provide advice that should help guide the user in selecting the appropriate instrumentation:

> A plan should be developed that addresses the following questions: which patients will be included in the study? What outcomes will be most

> effected by the treatment? When will the outcomes be measured? Who is going to read (and use) the information provided by the outcomes study? The more specific the answer to these questions, the better the choice of outcome instrument. (p. 4)

This and other recommendations would appear equally applicable when selecting instruments for other therapeutic assessment purposes.

One final set of criteria should be considered in the light of the following section on screening. Screening for the likelihood of the presence of disorders or for the need for additional assessment requires considerations that do not necessarily apply to instruments when they are used for the other therapeutic assessment purposes addressed in this chapter. A major one here is a specific consideration relative to a screener's criterion validity. Although broadly encompassed by the construct of "validity" that was previously discussed, it demands particular attention when evaluating instruments for screening purposes. What is being referred to here is the instrument's classification accuracy or efficiency.

Classification efficiency is usually expressed in terms of the following statistics: *sensitivity*, that is, the proportion of those individuals with the characteristic of interest who are accurately identified as such; *specificity*, that is, the proportion of individuals not having the characteristic of interest who are accurately identified as such); *positive predictive power*, which is the proportion of a population identified by the instrument as having the characteristic who actually do have the characteristic, and *negative predictive power*, which is the proportion of a population identified by the instrument as not having the characteristic who actually do not have the characteristic. This information can provide the clinical psychologist and other evaluators with empirically based information that is useful in the type of decision-making requiring the selection of one of two choices. The questions answered are typically those of the "yes/no" type, such as "Is the patient depressed or not?" or "Does the patient have a psychological problem significant enough to require treatment?" The reader is referred to Baldessarini, Finkelstein, and Arana (1983) for a discussion of issues related to the use of these statistics.

In evaluating these statistics, one must consider a few very important issues. One is the degree to which the clinician is willing to accept false-positives or false-negatives. This will be a function of the importance of maximizing the correct identification of those with the particular characteristic of interest vs. the importance of maximizing the correct identification of those not having the characteristic vs. the importance of optimizing the identification of both groups. This in turn will be dependent on the cutoff score recommended by the developer of the instrument and/or the efficiency values that are available when other cutoff scores are applied. These and related issues are discussed more extensively in the next section.

4.18.5 PSYCHOLOGICAL ASSESSMENT AS A TOOL FOR SCREENING

One of the most significant ways in which psychological assessment can contribute to the development of an economic and efficient behavioral health care delivery system is by using it to screen potential patients for need for behavioral health care services, and/or to determine the likelihood that the problem being screened is a particular disorder of interest. Probably the most concise, informative treatment of the topic of the use of psychological tests in screening for behavioral health care disorders is provided by Derogatis and Della-Pietra (1994). In this work, these authors turn to the Commission on Chronic Illness (1987) to provide a good working definition of health care screening in general, that being:

> the presumptive identification of unrecognized disease or defect by the application of tests, examinations or other procedures which can be applied rapidly to sort out apparently well persons who probably have a disease from those who probably do not. (Commission on Chronic Illness, 1987, p. 45)

Derogatis and DellaPietra (1994) further clarify the nature and the use of screening procedures, stating that:

> the screening process represents a relatively unrefined sieve that is designed to segregate the cohort under assessment into "positives" who presumptively have the condition, and "negatives" who are ostensibly free of the disorder. Screening is not a diagnostic procedure *per se*. Rather, it represents a preliminary filtering operation that identifies those individuals with the highest probability of having the disorder in question for subsequent specific diagnostic evaluation. Individuals found negative by the screening process are not evaluated further. (p. 23)

The most important aspect of any screening procedure is the efficiency with which it can provide information useful to clinical decision-making. In the area of clinical psychology, the most efficient and thoroughly investigated screening procedures involve the use of psychological assessment instruments. As implied by the foregoing, the power or utility of a psychological screener lies in its ability to determine, with a high level of probability, whether the respondent does or does not have a particular disorder or condition, or whether he or she is or is not a member of a group with clearly defined characteristics. In daily clinical practice, the most commonly used screeners are those designed specifically to identify some aspect of psychological functioning or disturbance or provide a broad overview of the respondent's point-in-time mental status. Examples of problem-specific screeners include the Beck Depression Inventory (BDI; Beck, Rush, Shaw, & Emery, 1979) and the State–Trait Anxiety Inventory (STAI; Spielberger, 1983). Examples of screeners for more generalized psychopathology or distress include the SA-45 and BSI.

4.18.5.1 Research-based Use of Psychological Screeners

The establishment of a system for screening for a particular disorder or condition involves determining what it is one wants to screen in or screen out, at what level of probability one feels comfortable in making that decision, and how many incorrect classifications or what percentage of errors one is willing to tolerate. Once it is decided what one wishes to screen for, one then must turn to the instrument's classification efficiency statistics, that is, sensitivity, specificity, positive predictive power (PPP), and negative predictive power (NPP), for the information necessary to determine whether a given instrument is suitable for the intended purpose(s).

Recall that sensitivity refers to the proportion of those with the characteristic of interest who are accurately identified as such by an instrument or procedure, while specificity refers to the proportion of those not having the characteristic of interest who are accurately identified. The cutoff score, index value, or other criterion used for classification can be adjusted to maximize either sensitivity or specificity. However, maximization of one will necessarily result in a decrease in the other, thus increasing the percentage of false-positives (with maximized sensitivity) or false-negatives (with maximized specificity). Stated differently, false-positives will increase as specificity decreases, while false-negatives will increase as sensitivity decreases (Elwood, 1993).

Another approach is to optimize both sensitivity and specificity, thus yielding a fairly even balance of true positives and true negatives. Although optimization might seem to be the preferable approach in all instances, there are situations in which a maximization approach is more desirable. For example, a psychiatric hospital with an inordinately high rate of inpatient suicide attempts begins to employ a screener designed to help identify patients with suicide potential as part of its admission procedures. The hospital adjusts the classification cutoff score to a level that identifies all suicidal patients in the screener's normative group. This cutoff score is then applied to all patients being admitted to the hospital for the purpose of identifying those requiring an extensive evaluation for suicide potential. This not only increases the number of true positives, but it also decreases the specificity and increases the number of false positives. However, the trade-off of identifying more suicidal patients early on with having more nonsuicidal patients receiving suicide evaluations would appear worthwhile for the hospital's purposes. Similarly, in other instances, maximization of specificity may be the preferred approach. For example, an MCO might wish to use a measure of overall level of psychological distress to identify those covered lives that are not in need of behavioral health care services. Sensitivity will decrease but, for the MCO's purposes, this might be quite acceptable.

Hsiao, Bartko, and Potter (1989) note that "a diagnostic test will not have a unique sensitivity and specificity. Instead, for each diagnostic test, the relationship between sensitivity and specificity depends on the cutoff point chosen for the test" (p. 665). The effect of employing individual classification cutoff points can be presented via the use of receiver operating characteristic (ROC) curves. These curves are nothing more than a plotting of the resulting true positive rate (sensitivity) against the false positive rate for each cutoff score that might be employed with a test used for classification purposes. The plotting allows for a graphical representation of what may be gained and/or lost by shifting cutoff scores. The resulting area underneath the curve provides an indication of how well the test performs. Development of ROC curves from available data for a test being considered for screening purposes is recommended. The reader is referred to Hsiao et al. and Metz (1978) for a more detailed discussion of ROC curves and their use.

In day-to-day clinical work, an instrument's PPP and NPP can provide information that is more useful than sensitivity and specificity. As Elwood (1993) has pointed out,

> Although sensitivity and specificity do provide important information about the overall performance of a test, their limitation in classifying

individual subjects becomes evident when they are considered in terms of conditional probabilities. Sensitivity is $P\ (+/d)$, the probability (P) of a positive test result ($+$) given that the subject has the target disorder (d). However, the task of the clinicians in assessing individual patients is just the opposite: determining $P\ (d/+)$, the probability that a patient has the disorder given that he or she obtained an abnormal test score. In the same way, specificity expresses $P\ (-/-d)$, the probability that a patient will have a negative test result given that he or she does not have the disorder. Here again, the task confronting the clinician is usually just the opposite: determining $P\ (-d/-)$, the probability that the patient does not have the disorder given a negative test result. (p. 410)

A note of caution is warranted when evaluating the two predictive powers of a test. Unlike sensitivity and specificity, both PPP and NPP are affected and change according to the prevalence or base rate at which the condition or characteristic of interest (i.e., that which is being screened by the test) occurs within a given setting. As Elwood (1993) reports, the lowering of base rates results in lower PPPs while increasing base rates result in higher PPPs. The opposite trend is true for NPPs. He notes that this is an important consideration because clinical tests are frequently validated using samples in which the prevalence rate is 0.50, or 50%. Thus, it is not surprising to see a test's PPP drop in "real-life" applications where the prevalence is lower.

Derogatis and DellaPietra (1994) indicate that a procedure referred to as "sequential screening" may provide at least a partial solution to the limitations or other problems that low base rates may pose for the predictive powers of an instrument. Sequential screening essentially involves the administration of two screeners, each of which measures the condition of interest, and two-phase screening. In the first phase, one screener is administered to the low base rate population. The purpose of this is to identify those individuals without the condition, thus requiring relatively good specificity. These individuals are eliminated from involvement in the second phase, resulting in an increase in the prevalence of the condition among those who remain. This group is then administered another screener of equal or better sensitivity. With the increased prevalence of the condition in the remaining group, the false positive rate will be much lower. As Derogatis and DellaPietra point out,

> Sequential screening essentially zeros in on a high-risk subgroup of the population of interest by virtue of a series of consecutive sieves. These have the effect of eliminating from consideration individuals with low likelihood of having the disorder, and simultaneously raising the base rate of the condition in the remaining sample. (p. 45)

In summary, PPP and NPP can provide information that is quite valuable to those making important clinical decisions, such as determining need for behavioral health care services, assigning diagnoses, or determining appropriate level of care. However, these users must be cognizant of the manner in which the predictive powers may change with the population to which the test or procedure is applied.

4.18.5.2 Implementation of Screeners into the Daily Work Flow of Service Delivery

The utility of a screening instrument is only as good as the degree to which it can be integrated into an organization's daily regimen of service delivery. This, in turn, depends on a number of factors. The first is the degree to which the administration and scoring of the screener is quick and easy, and the amount of time required to train the provider's staff to successfully incorporate the screener into their day-to-day activities.

The second factor relates to its use. Here, the screener is not used for anything other than determining the likelihood that the patient does or does not have the specific condition or characteristic the instrument is designed to assess. Use for any other purpose (e.g., assigning a diagnosis based solely on screener results, determining the likelihood of the presence of other characteristics) only serves to undermine the integrity of the instrument in the eyes of staff, payers, and other parties with a vested interest in the screening process.

The third factor has to do with the ability of the provider to act on the information. It must be clear how the clinician should proceed based on the information available.

The final factor is staff acceptance and commitment to the screening process. This comes only with a clear understanding of the importance of the screening, the usefulness of the obtained information, and how the screening process is to be incorporated into the organization's business flow.

Ficken (1995) provides an example of how screeners can be integrated into an assessment system designed to assist primary care physicians to identify patients with psychiatric disorders. This system (which also allows for the incorporation of practice guidelines) seems to take into account the first three utility-related factors listed above. It begins with the administration of a screener that is highly sensitive and specific to *DSM*- or *ICD*-related disorders. Ficken indicates that screeners should require no more than 10 minutes to complete, and that "their administration must be integrated seamlessly into the standard clinical routine" (p. 13). Somewhat similarly to the sequence described by Derogatis and DellaPietra (1994), positive findings would lead to a second level of testing. Here, another screener that meets the same requirements as those for the first screener and also affirms or rules out a diagnosis would be administered. Positive findings would lead to additional assessment for treatment planning purposes. Consistent with standard practice, Ficken recommends confirmation of screener findings by a qualified psychologist or physician.

4.18.6 PSYCHOLOGICAL ASSESSMENT AS A TOOL FOR TREATMENT PLANNING

The administration of screeners is only one way in which psychological assessment can serve as a valuable tool for treatment planning. However, many would argue that it is the most limited way in which this tool can be used for planning a course of treatment. When employed by a trained clinician, psychological assessment can provide information that can greatly facilitate and enhance the planning of the therapeutic intervention for the individual patient.

The importance of treatment planning has received significant attention during recent years. The reasons for this were summarized previously by this author (Maruish, 1990) as follows:

> Among important and interrelated reasons ... [are] concerted efforts to make psychotherapy more efficient and cost effective, the growing influence of "third parties" (insurance companies and the federal government) that are called upon to foot the bill for psychological as well as medical treatments, and society's disenchantment with open-ended forms of psychotherapy without clearly defined goals. (p. iii)

The role that psychological assessment can play in planning a course of treatment for behavioral health care problems is significant. Butcher (1990) indicated that information available from instruments such as the MMPI-2 can not only assist in identifying problems (see above) and establishing communication with the patient (see below), it can also help ensure that the plan for treatment is consistent with the patient's personality and external resources. In addition, psychological assessment may reveal potential obstacles to therapy, areas of potential growth, and problems of which the patient may not be consciously aware. Moreover, both Butcher and Appelbaum (1990) viewed testing as a means of quickly obtaining a second opinion. Other benefits of the results of psychological assessment, identified by Appelbaum, include assistance in identifying patient strengths and weaknesses, identification of the complexity of the patient's personality, and establishment of a reference point or guide to refer to during the therapeutic episode.

The types of information that can be derived from patient assessment and the manner in which it is applied for this purpose are quite varied—a fact that will become evident below. Nevertheless, Strupp (see Butcher, 1990) probably provided the best summary of the potential contribution of psychological assessment to treatment planning, stating that "careful assessment of patient's personality resources and liabilities is of inestimable importance. It will predictably save money and avoid misplaced therapeutic effort; it can also enhance the likelihood of favorable treatment outcomes for suitable patients" (pp. v–vi).

4.18.6.1 Assumptions About Treatment Planning

The introduction to this section presented a broad overview of ways in which psychological assessment can assist in devising and successfully implementing plans of treatment for behavioral health care patients. These and other benefits will be discussed in greater detail below. However, it is important to first clarify what treatment planning is and some of the general, implicit assumptions that one typically can make about this important therapeutic activity.

For the purpose of this discussion, the term "treatment planning" indicates that part of a therapeutic episode in which the treatment provider develops a set of goals for an individual presenting with behavioral health care problems, and outlines the specific means by which he/she or other resources will assist the patient in achieving those goals in the most efficient manner. General assumptions underlying the treatment planning process are as follows.

(i) The patient is experiencing behavioral health problems that have been identified either by themself or by another party. Common external sources of problem identification include the patient's spouse, parent, teacher, employer, and the legal system.

(ii) The patient experiences some degree of internal and/or external motivation to eliminate

or reduce the identified problems. An example of external motivation to change is the potential loss of job or marriage if problems are not resolved.

(iii) The goals of treatment are tied either directly or indirectly to the identified problems.

(iv) The goals of treatment have definable criteria for achievement, are indeed achievable by the patient, and are developed in collaboration with the patient.

(v) The prioritization of goals is reflected in the treatment plan.

(vi) The patient's progress toward achievement of the treatment goals can be tracked and compared against an expected path of improvement in either a formal or informal manner. This expected path of improvement may be based on the clinician's experience or (ideally) on objective data gathered on patients similar to the patient.

(vii) Deviations from the expected path of improvement will lead to a modification in the treatment plan, followed by subsequent monitoring to determine the effectiveness of the alteration.

These assumptions should not be considered exhaustive, nor are they reflective of what actually occurs in all situations. For example, some patients seen for therapeutic services may have no motivation to change. As may be seen in juvenile detention settings or in cases where children are brought to treatment by the parents, their participation in treatment is forced, and they may engage in intentional efforts to sabotage any therapeutic intervention. Also, it is likely that there are still clinicians who identify and prioritize treatment goals without the direct input of the patient. Nevertheless, the assumptions above represent this author's view of the aspects of treatment planning that have a direct bearing on the manner in which psychological assessment can best serve treatment planning efforts.

4.18.6.2 The Benefits of Psychological Assessment for Treatment Planning

As has already been touched upon, there are several ways in which psychological assessment can assist in the planning of treatment for behavioral health care patients. Following is a discussion of the more common and evident contributions that assessment can make to treatment planning efforts. These can be organized into four general categories: problem identification, problem clarification, identification of important patient characteristics, and monitoring of treatment progress.

4.18.6.2.1 Problem identification

Probably the most common use of psychological assessment in the service of treatment planning is for the purpose of problem identification. Often, the use of psychological testing *per se* is not needed to identify what problems the patient is experiencing. He or she will either tell the clinician directly without questioning, or they will readily admit to their problem(s) during the course of a clinical interview. However, this is not always the case.

The value of psychological testing becomes apparent in those cases where the patient is hesitant or unable to identify the nature of his or her problems. However, with a motivated and engaged patient who responds to items on a well validated and reliable test in an open and honest manner, the process of identifying what brought the patient to treatment also may be greatly facilitated. Cooperation shown during testing may be attributable to the nonthreatening nature of responding to questions presented on paper or a computer monitor (as opposed to those posed by another human being); the subtle, indirect, or otherwise nonthreatening nature of the questions (compared to those asked by the clinician); instrumentation that "casts a wider net" than the clinician in his or her interview with the patient; or any combination of these reasons.

In addition, the nature of some of the more commonly used psychological test instruments allows for the identification of secondary problems of significant severity that might otherwise be overlooked. Multidimensional inventories such as the MMPI-2 and the PAI are good examples of these types of instruments. Moreover, these instruments may be sensitive to other problems or patient traits or characteristics that may not necessarily be problems but which may exacerbate or otherwise contribute to the maintenance of the patient's problems.

Note that the type of problem identification described here is different from that conducted during screening (see above). Whereas screening is focused on determining the presence or absence of a single problem, problem identification generally takes a broader view and investigates the possibility of the presence of multiple problem areas. At the same time, there is also an attempt to determine the extent to which the problem area(s) affect the patient's ability to function.

4.18.6.2.2 Problem clarification

Psychological testing can often assist in the clarification of a known problem. Through tests designed for use with individuals presenting problems similar to the patient's, aspects of identified problems can be elucidated. This will improve the patient's and clinician's understanding of the problem and likely lead to a better treatment plan. The three most important types of information that can be gleaned for this purpose are the severity of the problems, the complexity of the problems, and the degree to which the problems impair the patient's ability to function in one or more life roles.

The manner in which a patient is treated depends a great deal on the severity of his or her problem. In particular, severity has a great bearing on the setting in which the behavioral health care intervention is provided. Those whose problems are so severe that they are considered a danger to themselves or others more often than not are best suited for inpatient treatment, at least until dangerousness is no longer an issue. Similarly, problem severity may be a primary criterion for an evaluation for a medication adjunct to treatment. Severity also may have a bearing on the type of psychotherapeutic approach that is taken by the clinician. For example, it may be more productive for the clinician to take a supportive role with severe cases; all things being equal, a more confrontational approach may be more appropriate with patients with problems in the mild to moderate range of severity.

As alluded to above, the problems of patients seeking behavioral health care services are frequently multidimensional. Patient and environmental factors that play into the formation and maintenance of a problem, along with the latter's relationship with other problems, all contribute to its complexity. Knowing the complexity of the target problems is invaluable in devising an effective treatment plan. Again, multidimensional instruments or batteries of tests measuring specific aspects of psychological dysfunction serve this purpose well.

As with problem severity, knowledge of the complexity of a patient's psychological problems can help the clinician and patient in many aspects of treatment planning, including determination of appropriate setting, therapeutic approach, need for medication, and other matters on which important decisions must be made. However, possibly of equal importance and concern to the patient and outside parties (spouse, employer, school, etc.) is the extent to which these problems affect the patient's ability to function in his or her role as parent, child, employee, student, friend, and so on. Data gathered from the administration of measures of role functioning can provide information that not only clarifies the impact of the patient's problems and serves to establish role-specific goals, but also identifies other parties that may serve as potential allies in the therapeutic process. In general, the most important role-functioning domains for assessment would be those related to work or school performance, interpersonal relationships, and activities of daily living (ADLs).

4.18.6.2.3 Identification of important patient characteristics

The identification and clarification of the patient's problems is of key importance in planning a course of treatment for the patient. However, there are numerous types of nonproblem-oriented patient information that can be useful in planning treatment and can be rather easily identified through the use of psychological assessment instruments. The vast majority of treatment plans are developed or modified with consideration of at least some of these other patient characteristics. The exceptions mostly are found with clinicians or programs that take a "one size fits all" approach to the treatment of general or specific types of disorders. It is beyond the scope of this chapter to provide an exhaustive list of what other types of information may be available to the clinician. However, a few are particularly worth mentioning.

Probably the most useful type of nonproblem-oriented information that can be gleaned from psychological assessment results is the identification of the patient characteristics or conditions that can serve as assets or areas of strength for the patient in working toward achieving the therapeutic goals. For example, Morey and Henry (1994) point to the utility of the PAI's Nonsupport scale in identifying whether the patient perceives an adequate social support network, this being a predictor of positive therapeutic progress. Other examples include "normal" personality characteristic information, such as that which can be obtained from Gough, McClosky, and Meehl's Dominance and Social Responsibility scales (1951, 1952) developed for use with the MMPI/MMPI-2. Greene (1991) indicates that those with high scores on the Dominance scale are described as "being able to take charge of responsibility for their lives. They are poised, self-assured, and confident of their own abilities" (p. 209). Gough and his colleagues interpreted high scores on the Social Responsibility scale as being indicative of individuals who, among other things, trust the world, are self-assured and poised, and stress the need for one to carry his or her share of duties. Thus, scores on these scales may reveal some important aspects of patient functioning that can be used in the service of affecting therapeutic change.

Similarly, knowledge of the patient's weaknesses or deficits may impact the type of treatment plan that is devised. Greene and Clopton (1994) provided numerous types of deficit-relevant information from the MMPI-2 Content Scales that have implications for treatment planning. For example, a clinically significant score ($T > 64$) on the Anger scale should lead one to consider the inclusion of training in assertiveness and/or anger control as part of the patient's treatment. On the other hand, uneasiness in social situations, as suggested by a significantly elevated score on either the Low Self-Esteem or Social Discomfort scale, suggests that a supportive approach to the intervention would be beneficial, at least initially.

Moreover, use of specially designed scales and procedures can provide information related to the patient's ability to become engaged in the therapeutic process. For example, the MMPI-2 Negative Treatment Indicators content scale developed by Butcher and his colleagues (Butcher, Graham, Williams, & Ben-Porath, 1989) may be useful in determining whether the patient is likely to be resistant to any form of "talk" therapy. Morey and Henry (1994) have supplied algorithms utilizing *T* scores for various PAI scales to make statements about the presence of positive characteristics, such as the presence of sufficient distress to motivate engagement in treatment, the ability to form a therapeutic alliance, and the capacity to utilize psychotherapy. The Therapeutic Reactance Scale (Dowd, Milne, & Wise, 1991) is yet another example of an instrument from which the clinician can be forewarned of potential resistance to therapeutic intervention.

Other types of patient characteristics that can be identified through psychological assessment have implications for the choice of the therapeutic approach and thus can contribute significantly to the treatment planning process. Beutler and his colleagues (Beutler & Clarkin, 1990; Beutler, Wakefield, & Williams, 1994; Beutler & Williams, 1995) have identified four patient characteristics that are thought to be important to matching patients and treatment approach for maximized therapeutic effectiveness. These include symptom severity, symptom complexity, coping style, and potential resistance to treatment. At different points in time, other patient variables also have been identified by these investigators as important considerations in the selection of the best treatment for a given patient. These include the problem-solving phase the patient has reached (Beutler & Clarkin, 1990), and subjective distress and social support (L.E. Beutler, personal communication, January 15, 1996).

Moreland (1996) points out how psychological assessment can assist in determining whether the patient deals with problems through internalizing or externalizing behaviors. All things being equal, internalizers would probably profit most from an insight-oriented approach rather than a behaviorally oriented approach. The reverse would be true for externalizers. In addition, cognitive factors also are important. Knowing that intelligence test results indicate an average or above IQ can assist the clinician in determining whether a patient will be able to benefit from a cognitive approach.

4.18.6.2.4 Monitoring of progress along the path of expected improvement

Information from repeated testing during the treatment process can help the clinician to determine if the treatment plan is appropriate for the patient at that particular point in time. Thus, many clinicians use psychological assessment to determine whether their patients are showing the expected improvement as treatment progresses. If not, adjustments can be made. These adjustments may reflect the need for a more intensive or aggressive treatment approach (e.g., increased number of psychotherapeutic sessions each week, addition of a medication adjunct) or for a less intensive approach (e.g., reduce or terminate medication, transfer from inpatient to outpatient care). Either way, this may require further retesting in order to determine whether the treatment revisions have impacted the course of change in the expected direction. This process may be repeated any number of times. In-treatment retestings also can provide information relevant to the decision of when to terminate treatment.

The goal of monitoring is to determine whether treatment is "on track" with the progress that is expected at a given point in time. When and how often one might assess the patient is dependent on a few factors. The first is the instrumentation. Many instruments are designed to assess the patient's status at the time of testing. Items on these measures are generally worded in the present tense (e.g., "I feel tense and nervous," "I feel that my family loves and cares about me"). Changes from one day to the next on the constructs measured by the instrument should be reflected in the test results.

Other instruments, however, ask the patient to indicate if a variable of interest has been present, or how much or to what extent it has occurred during a specific time period in the past. The items usually are asked in the context of something like "During the past month, how often have you ..." or "During the past week, to what extent has ..." Readministration of a measure containing interval-of-time-specific items or subsets of items should be undertaken only after a period of time equivalent to or longer than the time interval to be considered in responding to the items has past. For example, an instrument which asks the patient to consider how much certain symptoms have been problematic during the past seven days should not be readministered for at least seven days. The responses elicited during a readministration that occurs less than seven days after the first administration would include the patient's consideration of his or her status during the previously considered time period. This may make interpretation of the change of symptom status (if any) from the first to the second assessment difficult if not impossible.

Methods to determine whether clinically significant change has occurred from one point in time to another have been developed and can be used for treatment monitoring purposes. These are discussed in the outcomes assessment section of this chapter below. However, for monitoring purposes, another approach to evaluating therapeutic change may be superior. This approach may be referred to as the "glide path" approach, with the term referring to the narrow descent course or path that airplanes must follow when landing. Deviation from the flight glide path requires corrections in the plane's speed, altitude, and/or attitude in order to return to the glide path and a safe landing.

R.L. Kane (personal communication, July 22, 1996) has indicated that just as a pilot has the instrumentation to alert him or her about the plane's position on the glide path, the clinician may use psychological assessment instruments to track how well the patient is following the glide path of treatment. The glide path in this case represents expected improvement over time in one or more measurable areas of functioning (e.g., symptom severity, social role functioning, occupational performance). The expectations would be based on objective data obtained from similar patients at various points during their treatment and would allow for minor deviations from the path. The end of the glide path is one or more specific goals that are part of the treatment plan. Thus, "arrival" at the end of the glide path signifies the attainment of specific treatment goals.

4.18.7 PSYCHOLOGICAL ASSESSMENT AS A THERAPEUTIC INTERVENTION

The use of psychological assessment as a means of therapeutic intervention in and of itself has received more than passing attention during the past few years. "Therapeutic assessment" with the MMPI-2 has received particular attention primarily through the work of Finn and his associates (Finn, 1996a, 1996b; Finn & Martin, in press; Finn & Tonsager, 1992). Finn's approach appears to be applicable with instruments or batteries of instruments that provide multidimensional information relevant to the concerns of patients seeking answers to questions related to their mental health status. The approach espoused by Finn thus will be presented here as a model for deriving direct therapeutic benefits from the psychological assessment experience.

4.18.7.1 What Is Therapeutic Assessment?

In discussing the use of the MMPI-2 as a therapeutic intervention, Finn (1996a) describes an assessment procedure whose goal is to "gather accurate information about clients ... and then use this information to help clients understand themselves and make positive changes in their lives" (p. 3). Elaborating on this procedure and extending it to the use of any test, Finn and Martin (in press) describe therapeutic assessment as

> collaborative, interpersonal, focused, time limited, and flexible. It is ... very interactive and requires the greatest of clinical skills in a challenging role for the clinician. It is unsurpassed in a respectfulness for clients: collaborating with them to address *their* concerns (around which the work revolves), acknowledging them as experts on themselves and recognizing their contributions as essential, and providing to them usable answers to their questions in a therapeutic manner.

The ultimate goal of therapeutic assessment is to provide an experience for the client that will allow him/her to take steps toward greater psychological health and a more fulfilling life. This is done by recognizing the client's characteristic ways of being, understanding in a meaningful, idiographic way the problems the client faces, providing a safe environment for the client to explore change, and providing the opportunity for the client to experience new ways of being in a supportive environment.

Simply stated, therapeutic assessment may be considered an approach to the assessment of mental health patients in which the patient is not only the primary provider of information needed to answer questions, but also is actively involved in formulating the questions that are to be answered by the assessment. Feedback regarding the results of the assessment is provided to the patient and is considered a

Similarly, knowledge of the patient's weaknesses or deficits may impact the type of treatment plan that is devised. Greene and Clopton (1994) provided numerous types of deficit-relevant information from the MMPI-2 Content Scales that have implications for treatment planning. For example, a clinically significant score ($T > 64$) on the Anger scale should lead one to consider the inclusion of training in assertiveness and/or anger control as part of the patient's treatment. On the other hand, uneasiness in social situations, as suggested by a significantly elevated score on either the Low Self-Esteem or Social Discomfort scale, suggests that a supportive approach to the intervention would be beneficial, at least initially.

Moreover, use of specially designed scales and procedures can provide information related to the patient's ability to become engaged in the therapeutic process. For example, the MMPI-2 Negative Treatment Indicators content scale developed by Butcher and his colleagues (Butcher, Graham, Williams, & Ben-Porath, 1989) may be useful in determining whether the patient is likely to be resistant to any form of "talk" therapy. Morey and Henry (1994) have supplied algorithms utilizing T scores for various PAI scales to make statements about the presence of positive characteristics, such as the presence of sufficient distress to motivate engagement in treatment, the ability to form a therapeutic alliance, and the capacity to utilize psychotherapy. The Therapeutic Reactance Scale (Dowd, Milne, & Wise, 1991) is yet another example of an instrument from which the clinician can be forewarned of potential resistance to therapeutic intervention.

Other types of patient characteristics that can be identified through psychological assessment have implications for the choice of the therapeutic approach and thus can contribute significantly to the treatment planning process. Beutler and his colleagues (Beutler & Clarkin, 1990; Beutler, Wakefield, & Williams, 1994; Beutler & Williams, 1995) have identified four patient characteristics that are thought to be important to matching patients and treatment approach for maximized therapeutic effectiveness. These include symptom severity, symptom complexity, coping style, and potential resistance to treatment. At different points in time, other patient variables also have been identified by these investigators as important considerations in the selection of the best treatment for a given patient. These include the problem-solving phase the patient has reached (Beutler & Clarkin, 1990), and subjective distress and social support (L.E. Beutler, personal communication, January 15, 1996).

Moreland (1996) points out how psychological assessment can assist in determining whether the patient deals with problems through internalizing or externalizing behaviors. All things being equal, internalizers would probably profit most from an insight-oriented approach rather than a behaviorally oriented approach. The reverse would be true for externalizers. In addition, cognitive factors also are important. Knowing that intelligence test results indicate an average or above IQ can assist the clinician in determining whether a patient will be able to benefit from a cognitive approach.

4.18.6.2.4 Monitoring of progress along the path of expected improvement

Information from repeated testing during the treatment process can help the clinician to determine if the treatment plan is appropriate for the patient at that particular point in time. Thus, many clinicians use psychological assessment to determine whether their patients are showing the expected improvement as treatment progresses. If not, adjustments can be made. These adjustments may reflect the need for a more intensive or aggressive treatment approach (e.g., increased number of psychotherapeutic sessions each week, addition of a medication adjunct) or for a less intensive approach (e.g., reduce or terminate medication, transfer from inpatient to outpatient care). Either way, this may require further retesting in order to determine whether the treatment revisions have impacted the course of change in the expected direction. This process may be repeated any number of times. In-treatment retestings also can provide information relevant to the decision of when to terminate treatment.

The goal of monitoring is to determine whether treatment is "on track" with the progress that is expected at a given point in time. When and how often one might assess the patient is dependent on a few factors. The first is the instrumentation. Many instruments are designed to assess the patient's status at the time of testing. Items on these measures are generally worded in the present tense (e.g., "I feel tense and nervous," "I feel that my family loves and cares about me"). Changes from one day to the next on the constructs measured by the instrument should be reflected in the test results.

Other instruments, however, ask the patient to indicate if a variable of interest has been present, or how much or to what extent it has occurred during a specific time period in the past. The items usually are asked in the context of something like "During the past month, how often have you . . ." or "During the past week, to what extent has . . ." Readministration of a measure containing interval-of-time-specific items or subsets of items should be undertaken only after a period of time equivalent to or longer than the time interval to be considered in responding to the items has past. For example, an instrument which asks the patient to consider how much certain symptoms have been problematic during the past seven days should not be readministered for at least seven days. The responses elicited during a readministration that occurs less than seven days after the first administration would include the patient's consideration of his or her status during the previously considered time period. This may make interpretation of the change of symptom status (if any) from the first to the second assessment difficult if not impossible.

Methods to determine whether clinically significant change has occurred from one point in time to another have been developed and can be used for treatment monitoring purposes. These are discussed in the outcomes assessment section of this chapter below. However, for monitoring purposes, another approach to evaluating therapeutic change may be superior. This approach may be referred to as the "glide path" approach, with the term referring to the narrow descent course or path that airplanes must follow when landing. Deviation from the flight glide path requires corrections in the plane's speed, altitude, and/or attitude in order to return to the glide path and a safe landing.

R.L. Kane (personal communication, July 22, 1996) has indicated that just as a pilot has the instrumentation to alert him or her about the plane's position on the glide path, the clinician may use psychological assessment instruments to track how well the patient is following the glide path of treatment. The glide path in this case represents expected improvement over time in one or more measurable areas of functioning (e.g., symptom severity, social role functioning, occupational performance). The expectations would be based on objective data obtained from similar patients at various points during their treatment and would allow for minor deviations from the path. The end of the glide path is one or more specific goals that are part of the treatment plan. Thus, "arrival" at the end of the glide path signifies the attainment of specific treatment goals.

4.18.7 PSYCHOLOGICAL ASSESSMENT AS A THERAPEUTIC INTERVENTION

The use of psychological assessment as a means of therapeutic intervention in and of itself has received more than passing attention during the past few years. "Therapeutic assessment" with the MMPI-2 has received particular attention primarily through the work of Finn and his associates (Finn, 1996a, 1996b; Finn & Martin, in press; Finn & Tonsager, 1992). Finn's approach appears to be applicable with instruments or batteries of instruments that provide multidimensional information relevant to the concerns of patients seeking answers to questions related to their mental health status. The approach espoused by Finn thus will be presented here as a model for deriving direct therapeutic benefits from the psychological assessment experience.

4.18.7.1 What Is Therapeutic Assessment?

In discussing the use of the MMPI-2 as a therapeutic intervention, Finn (1996a) describes an assessment procedure whose goal is to "gather accurate information about clients . . . and then use this information to help clients understand themselves and make positive changes in their lives" (p. 3). Elaborating on this procedure and extending it to the use of any test, Finn and Martin (in press) describe therapeutic assessment as

> collaborative, interpersonal, focused, time limited, and flexible. It is . . . very interactive and requires the greatest of clinical skills in a challenging role for the clinician. It is unsurpassed in a respectfulness for clients: collaborating with them to address *their* concerns (around which the work revolves), acknowledging them as experts on themselves and recognizing their contributions as essential, and providing to them usable answers to their questions in a therapeutic manner.

The ultimate goal of therapeutic assessment is to provide an experience for the client that will allow him/her to take steps toward greater psychological health and a more fulfilling life. This is done by recognizing the client's characteristic ways of being, understanding in a meaningful, idiographic way the problems the client faces, providing a safe environment for the client to explore change, and providing the opportunity for the client to experience new ways of being in a supportive environment.

Simply stated, therapeutic assessment may be considered an approach to the assessment of mental health patients in which the patient is not only the primary provider of information needed to answer questions, but also is actively involved in formulating the questions that are to be answered by the assessment. Feedback regarding the results of the assessment is provided to the patient and is considered a

primary, and possibly the principal element of the assessment process. Thus, the patient becomes a partner in the assessment process; as a result, therapeutic and other benefits accrue.

The reader should note that in this section, the term "therapeutic assessment" is used to denote the specific approach advocated by Finn and his colleagues for using the psychological assessment process as an opportunity for therapeutic intervention. It should not be confused with the more general term "therapeutic psychological assessment" as it has been employed throughout this chapter; "therapeutic assessment" is but one aspect of therapeutic psychological assessment.

4.18.7.2 The Impetus for Therapeutic Assessment

To say that clinical psychologists performing psychological assessments in mental health settings traditionally have never shared much of their findings with their patients is probably not an overstatement. A common scenario throughout many mental health settings was that a patient being treated by a psychologist was evaluated by the latter, or a patient was referred to the psychologist by another mental health professional for assessment only. In the first instance, the degree to which the psychologist might directly share the results of the often lengthy and expensive evaluation would vary. Generally, a detailed review of the findings would be a rarity. In the latter instance, the patient would be evaluated, a report of the results dictated, and a copy of the report sent back to the referring clinician. In either instance, the purpose of the assessment probably would be to answer questions posed by the treating clinician. Unfortunately, the patient and his or her concerns as they related to the psychological assessment were typically of only secondary consideration, if any.

Fortunately, recent occurrences have begun to change the way in which assessment information is used. Consequently, the degree to which the patient is involved in the assessment process is changing. One reason for this is the relatively recent revision of the ethical standards of the APA (1992). This revision included a mandate for psychologists to provide feedback to clients whom they test. According to ethical standard 2.09:

> Unless the nature of the relationship is clearly explained to the person being assessed in advance and precludes provision of an explanation of results (such as in some organizational consulting, preemployment or security screenings, and forensic evaluations), psychologists ensure that an explanation of the results is provided using language that is reasonably understandable to the person assessed or to another legally authorized person on behalf of the client. Regardless of whether the scoring and interpretation are done by the psychologist, by assistants, or by automated or other outside services, psychologists take reasonable steps to ensure that appropriate explanations of results are given. (p. 8)

Many clinicians and other psychologists involved in assessment activities (e.g., counseling psychologists, neuropsychologists) have had to modify their practice routine to accommodate this requirement. Some view this requirement as resulting in an improvement in the quality of their services; others likely see it as nothing more than an inconvenience which, in the era of managed care and limited access to treatment, further limits the amount of time they have to work with a patient. However, most would agree that the patient has benefited from the required feedback.

Finn and Tonsager (1992) identified other factors that may have contributed to the recent interest in providing patients with assessment feedback. One is another external influence, that is, the recognition of the patient's right to see their medical and psychiatric health care records. However, they also point to several clinically and research-based findings and impressions that suggest that therapeutic assessment enhances patient care through the facilitation of patient–therapist rapport, cooperation during the assessment process, positive feelings about the process and the clinician, improvement in mental health status, and feelings of being understood by another. In addition, Finn and Tonsager refer to Finn and Butcher's (1991) summary of potential benefits that may accrue from providing test results feedback. The listed benefits, based on clinical experience, include increased feelings of self-esteem and hope, reduced symptomatology and feelings of isolation, increased understanding and self-awareness, and increased motivation to seek or be more actively involved in mental health treatment. Finally, Finn and Martin (in press) note that the therapeutic assessment process can lead to increased feelings of mastery and control and decreased feelings of alienation. At the same time, it can serve as a model for relationships that can result in mutual respect and the patient being seen for who he or she is.

4.18.7.3 The Therapeutic Assessment Process

Finn (1996a) has outlined a three-step procedure for therapeutic assessment using the MMPI-2. As indicated above, it should work equally well with other multidimensional instruments that one might select. Finn describes this procedure as one to be used in those situations in which the patient is seen only for assessment (i.e., the patient is not to be treated later by the assessing clinician). From the present author's standpoint, the procedures are equally applicable for use by clinicians who test patients whom they later treat. With these points in mind, the three-step procedure is summarized below.

4.18.7.3.1 Step 1: The initial interview

According to Finn (1996a), the initial interview with the patient serves multiple purposes. It provides an opportunity to build rapport, or to increase rapport if a patient–therapist relationship already exists. The assessment task is presented as a collaborative one, and the patient is given the opportunity to identify questions that he or she would like answered using the assessment data. Background information related to the patient-identified questions is subsequently gathered. Any reservations about participating in the therapeutic assessment process (e.g., confidentiality, previous negative experiences with assessment) are dealt with in order to facilitate maximal involvement in the process.

After responding to the patient's concerns, Finn (1996a) recommends that the clinician restate the questions posed earlier by the patient. This ensures the accuracy of what the patient would like to have addressed by the assessment. The patient also is encouraged to ask questions of the clinician, thus reinforcing the collaborative context or atmosphere that the clinician is trying to establish. Step 1 is completed as the instrumentation and its administration, as well as the responsibilities and expectations of each party, are clearly defined and the particulars of the process (e.g., date and time of assessment, date and time of the feedback session, clinician fees) are discussed and agreed upon.

4.18.7.3.2 Step 2: Preparing for the feedback session

Upon completion of the administration and scoring of the instrumentation used during the assessment, the clinician first outlines all results obtained from the assessment, including those not directly related to the patient's previously stated questions. Finn (1996a) presents a well-organized outline for the types of information that the trained user can extract from MMPI-2 data. These include response consistency, test-taking attitude, distress and disturbance, major symptoms, underlying personality, behavior in relationships, implications for treatment, diagnostic impression, and recommendations. Unfortunately, clinicians who do not or cannot use the MMPI-2 or other well-researched, multidimensional instruments will not have the same amount or type of data available to them. (This should not preclude them from identifying the types of valid and useful information that can be derived from the instruments and organizing it into a usable form for presentation to the patient.) This is followed by a determination of how to present the results to the patient. This can be guided by the clinician asking himself or herself the following questions:

> (i) How do the (test) findings relate to the client's goals?
> (ii) What are the most important findings of the (tests administered)?
> (iii) To what extent is the client likely to already know about and agree with the (test) findings?
> (iv) How much new information is the client likely to be able to integrate in the feedback session?
> (v) What is likely to happen if the client becomes overwhelmed or is presented with findings that are greatly discrepant from his/her current self-concept? (p. 34)

As a final point in this step, Finn (1996a) indicates that the clinician must determine what is the best way to present the information to the patient so that he or she can accept and integrate the information while maintaining his or her sense of identity and self-esteem. This also is a time when the clinician can identify information that he or she may not wish to reveal to the patient because it is not important to answering the patient's questions; doing so may negatively affect the collaborative relationship. In addition, the clinician may want to prepare for presenting those aspects of feedback that he or she feels will be most problematic for him or her (i.e., the clinician) by role-playing with a colleague.

4.18.7.3.3 Step 3: The feedback session

As Finn (1996a) states: "The overriding goal of feedback sessions is to have a therapeutic interaction with clients" (p. 44). Thus, the initial tasks of the feedback session are focused on setting the stage for this type of encounter. This is accomplished by allaying any anxiety the patient may have about the session, reaffirming the collaborative relationship, and familiarizing him or her with the presentation of the test results (e.g., explaining the profile sheet upon which the results are graphed, discussing the normative group to which he or she will be compared, providing an explanation of standard scores).

When the session preparation is completed, the clinician begins providing feedback to the patient (Finn, 1996a). This, of course, is centered on answering the questions posed by the patient during Step 1. Beginning with a positive finding from the assessment, the clinician proceeds to first address those questions that the patient is most likely to accept. He or she then carefully moves to the findings that are more likely to be anxiety-arousing for the patient and/or challenge his or her self-concept. A key element to this step is to have the patient verify the accuracy of each finding and provide a real-life example of the interpretation that is offered. Alternately, one should ask the patient to modify the interpretation to make it more in line with how he or she sees themselves and their situation. Finn (1996a) provides specific suggestions about how to deal with a rejection of a finding, the final suggestion being to allow the client to disagree with but not totally dismiss the finding. This leaves the door open for representing the finding at another time when the patient is more open to accepting it.

Finn (1996a) recommends that the clinician should end the session by responding to any additional questions the patient may have; confirming that the patient has accurately understood the information that was presented; giving permission for the patient to contact the clinician should further questions arise; and (in the assessment-only arrangement) termination of the relationship. Throughout the session, the clinician maintains a supportive stance with regard to any affective reactions to the findings.

4.18.7.3.4 Additional steps

Finn and Martin (in press) indicate two additional steps that may be added to the therapeutic assessment process. The purpose of the first additional step, referred to as an "assessment intervention session" essentially is to clarify initial test findings through the administration of additional instruments. For example, Finn and Martin explain how MMPI-2 findings can be further fleshed out through a nonstandard administration of an instrument such as the Thematic Apperception Test (TAT). Here, the clinician controls the patient's interpretation in order to draw out information relevant to the patient's questions. Also, solutions to problems elicited by the TAT cards are suggested to the patient.

The other additional step discussed by Finn and Martin (in press) is the provision of a written report of the findings to the patient. In addition to summarizing both the test results and the answers to the patient's questions, it also attempts to elicit feedback and reactions from the patient about the assessment.

The reader should note that the preceding summary presents only the key technical aspects of the therapeutic assessment procedures espoused by Finn and his associates. Much of the clinical/dynamic aspect of this approach has not been addressed because of the focus of this chapter. Those interested in incorporating the process into their clinical practice are encouraged to read Finn (1996a).

4.18.7.4 Empirical Support for Therapeutic Assessment

Noting the lack of direct empirical support for the therapeutic effects of sharing test results with patients, Finn and Tonsager (1992) investigated the benefits of providing feedback to university counseling center clients regarding their MMPI-2 results. A total of 32 subjects underwent therapeutic assessment and feedback procedures similar to those described above while on the counseling center's waiting list. Another 28 subjects were recruited from the same waiting list to serve as a control group. There were no significant differences between the two groups on any important demographic or examiner contact-interval variables.

Instead of receiving feedback, Finn and Tonsager's (1992) control group received nontherapeutic attention from the examiner. However, they were administered the same dependent measures as the feedback group at the same time as the experimental group received feedback. They also were administered the same dependent measures as the experimental group two weeks later (i.e., two weeks after the experimental group received the feedback) in order to determine if there were differences between the two groups on those dependent measures. These measures included a self-esteem questionnaire, a symptom checklist (i.e., the SCL-90-R), a measure of private and public self-consciousness, and a questionnaire assessing the subjects' subjective impressions of the feedback session. (Note that the control group was administered only that portion of the feedback assessment questionnaire that was relevant to them.)

The results of Finn and Tonsager's (1992) study indicated that compared to the control group, the feedback group demonstrated significantly less distress at the two-week post-feedback follow up, and significantly higher levels of self-esteem and hope at both the time of feedback and the two-week post-feedback follow up. In other findings, feelings about the feedback sessions were positively and significantly correlated with changes in self-esteem from testing to feedback, both from feedback to follow up and from testing to follow up among those who were administered the MMPI-2. In addition, the change in level of distress from feedback to follow up correlated significantly with private self-consciousness (i.e., the tendency to focus on the internal aspects of oneself) but not with public self-consciousness.

4.18.8 PSYCHOLOGICAL ASSESSMENT AS A TOOL FOR OUTCOMES MANAGEMENT

The 1990s have witnessed a positively accelerating growth curve reflecting the level of interest in and development of behavioral health care outcomes programs. Cagney and Woods (1994) attribute this to four major factors. First, behavioral health care purchasers are asking for information regarding the value of the services they buy. Second, there is an increasing number of purchasers who are requiring a demonstration of patient improvement and satisfaction. Third, MCOs need data that demonstrate that their providers render efficient and effective services. And fourth, outcomes information will be needed for the "quality report cards" that MCOs anticipate they will be required to provide in the future. In short, fueled by soaring health care costs, there has been an increasing need for providers to demonstrate that what they do is effective. And all of this has occurred within the context of the continuous quality improvement (CQI) movement, in which there have been similar trends in the level of interest and growth.

As this author has noted previously, the interest in and necessity for outcomes measurement in the era of managed care and accountability provides a unique opportunity for clinical psychologists to use their training and skills in assessment (Maruish, 1994). However, the extent to which the clinical psychologist becomes a key and successful contributor to an organization's outcomes initiative (whatever that might be) will depend on his or her understanding of what "outcomes" and their measurement and applications are all about.

4.18.8.1 What Are Outcomes?

Before discussing outcomes, it is important to have a clear understanding of what is meant by the term. Experience has shown that the meaning varies according to whom one may speak.

Donabedian (1985) has identified three dimensions of quality of care. "Structure" refers to the organization providing the care. It includes aspects such as how the organization is "organized," the physical facilities and equipment, and the number and professional qualifications of its staff. "Process" refers to the specific types of services that are provided to a given patient (or group of patients) during a specific episode of care. These might include various types of tests and assessments (e.g., psychological tests, lab tests, magnetic resonance imaging), therapeutic interventions (e.g., group psychotherapy, medication), and discharge planning activities. Treatment complications (e.g., drug reactions) are also included here. "Outcomes", on the other hand, refers to the results of the specific treatment that was rendered.

The outcomes, or results, of treatment should not refer to change in only a single aspect of functioning. Treatment may impact various facets of a patient's life. Stewart and Ware (1992) have identified five broad aspects of general health status: physical health, mental health, social functioning, role functioning, and general health perception. Treatment may affect these aspects of health in different ways, depending on the disease or disorder being treated and the effectiveness of the treatment. Some specific aspects of functioning related to these five areas of general health status that are commonly measured include: feeling of well-being, psychological symptom status, use of alcohol and other drugs, functioning on the job or at school, marital/family relationships, utilization of health care services, and ability to cope.

In considering the various types of outcomes that might be assessed in behavioral health care settings, a substantial number of clinicians probably would identify symptomatic change in psychological status as being the most important. Nevertheless, however important change in symptom status may have been in the past, clinical psychologists and other behavioral health care providers have come to realize that changes in many of the other aspects of functioning identified by Stewart and Ware (1992) are equally important indicators of treatment effectiveness. As Sederer et al. (1996) have noted:

> Outcome for patients, families, employers, and payers is not simply confined to symptomatic change. Equally important to those affected by the care rendered is the patient's capacity to function within a family, community, or work environment or to exist independently, without undue burden to the family and social welfare

> system. Also important is the patient's ability to show improvement in any concurrent medical and psychiatric disorder ... Finally, not only do patients seek symptomatic improvement, but they want to experience a subjective sense of health and well being. (p. 2)

A much broader perspective is offered in Faulker and Gray's *The 1995 behavioral outcomes and guidelines sourcebook* (Migdail, Youngs, & Bengen-Seltzer, 1995):

> Outcomes measures are being redefined from a vague "is the patient doing better?" to more specific questions, such as, "Does treatment work in ways that are measurably valuable to the patient in terms of daily functioning level and satisfaction, to the payor in terms of value for each dollar spent, to the managed care organization charged with administering the purchaser's dollars, and to the clinician charged with demonstrating value for hours spent?" (p. 1)

Thus, "outcomes" holds a different meaning for each of the different parties who have a stake in behavioral health care delivery. What is measured generally depends on the purposes for which outcomes assessment is undertaken. As will be shown, these vary greatly.

4.18.8.2 Outcomes Assessment: Measurement, Monitoring, and Management

Just as it is important to be clear about what is meant by outcomes, it is equally important to clarify the three general purposes for which outcomes assessment may be employed. The first is outcomes measurement. This involves nothing more than pre- and post-treatment assessment of one or more variables to determine the amount of change that has occurred (if any) in these variables as a result of therapeutic intervention.

A more useful approach is that of outcomes monitoring. This refers to "the use of periodic assessment of treatment outcomes to permit inferences about what has produced change" (Dorwart, 1996, p. 46). Like treatment progress monitoring used for treatment planning purposes, outcomes monitoring involves the tracking of changes in the status of one or more outcomes variables at multiple points in time. Assuming a baseline assessment at the beginning of treatment, reassessment may occur one or more times during the course of treatment (e.g., weekly, monthly), at the time of termination, and/or during one or more periods of post-termination follow up. Whereas treatment progress monitoring is used to determine how much the patient is on or off the expected course of improvement, outcomes monitoring focuses on revealing aspects about the therapeutic process that seem to affect change.

The third and most useful purpose of outcomes assessment is that of outcomes management. Dorwart (1996) defines outcomes management as "the use of monitoring information in the management of patients to improve both the clinical and administrative processes for delivering care" (pp. 46–47). In outcomes management, information is used to improve the quality of services offered to the patient population(s) served by the provider, not to any one patient. Information gained through the assessment of patients can provide the organization with indications of what works best with whom and under what set of circumstances, thus helping to improve the quality of services for all patients. In essence, outcomes management can serve as a tool for those organizations with an interest in implementing a CQI initiative.

4.18.8.3 The Benefits of Outcomes Assessment

The implementation of any type of outcomes assessment initiative within an organization does not come without effort from and cost to the organization. However, if it is implemented properly, all interested parties, that is, patients, clinicians, provider organizations, payers, and the health care industry as a whole, should find a substantial yield from the outlay of time and money. Cagney and Woods (1994) identify several benefits to patients, including enhanced health and quality of life, improved health care quality, and effective use of the dollars paid into benefits plans. For providers, the outcomes data can result in improved clinical skills, information related to the quality of care provided and local practice standards, increased profitability, and decreased concerns over possible litigation.

Outside of the clinical context, benefits also can accrue to payers and MCOs. Cagney and Woods (1994) see the potential payer benefits as including healthier workers, improved health care quality and worker productivity, and reduced or contained health care costs. As for MCOs, the benefits include increased profits, information that can help shape the practice patterns of their providers, and decisions that are based on quality of care.

4.18.8.4 The Therapeutic Use of Outcomes Assessment

The foregoing overview of outcomes assessment provides the background necessary for discussing the use of psychological outcomes assessment data in day-to-day clinical practice.

Whereas the focus of the above review was centered on both the individual patient and patient populations, it now will narrow to the use of outcomes assessment primarily in service to the individual patient. The reader interested in issues related to large, organization-wide outcomes studies conducted for outcomes management purposes (as defined above) is encouraged to seek other sources of information that specifically address that topic (see, for example, Migdail, Youngs, & Bengden-Seltzer, 1995; Newman, 1994).

There is no one system or approach to the assessment of treatment outcomes for an individual patient that is appropriate for all providers of behavioral health care services. Because of the specific type of outcomes one is interested in, the reasons for assessing them, and the manner in which they may impact the decisions made by the patient, payer and clinician, any successful and useful outcomes assessment approach must be customized. Customization should reflect the needs of the primary beneficiary of the information gained from the assessment (i.e., patient, payer, or provider), with consideration of the secondary stakeholders in the therapeutic endeavor. Ideally, the identified primary beneficiary would be the patient. Although this is not always the case, it would appear that only rarely would the patient not benefit, at least indirectly, from the gathering of outcomes data.

Following are considerations and recommendations for the development and implementation of an outcomes assessment initiative by behavioral health care providers. Although space limitations do not allow a comprehensive review of all issues and solutions, the information that follows can be useful to clinical psychologists and others with similar training wishing to begin to incorporate outcomes assessment into their standard therapeutic routine.

4.18.8.4.1 Purpose of the outcomes assessment

There are numerous reasons for assessing outcomes. For example, in a recent survey of 73 behavioral health care organizations, various reasons were identified by the participants as to why they had conducted outcomes studies (Pallak, 1994). Among the several indicated, the top five reasons (in descending order) were to: evaluate outcomes for patients, evaluate provider effectiveness, evaluate integrated treatment programs, manage individual patients, and support sales and marketing efforts. However, from the clinician's standpoint, a couple of purposes are worth noting.

In addition to monitoring the course of progress during treatment (see above), clinicians may employ outcomes assessment to obtain a direct measure of how much patient improvement has occurred as the result of the course of treatment intervention. Here, the findings are of more benefit to the clinician than to the patient himself because a pre- and post-treatment approach to the assessment is utilized. The information will not lead to any change in the patient providing the information, but the feedback it provides to the clinician could assist him in the treatment of other patients later on.

Another common reason for outcomes assessment is to demonstrate the patient's need for therapeutic services beyond that which is typically covered by the patient's medical and behavioral health care benefits. When assessment is conducted for this reason, the patient and clinician are only secondary beneficiaries of the outcomes data. As will be shown below, the type of information that a third party payer requires for authorization of extended benefits may not be the most relevant or useful to the patient or the clinician.

4.18.8.4.2 What to measure

The aspects or dimensions of patient functioning that are measured as part of outcomes assessment will depend on the purpose for which the assessment is being conducted. As discussed earlier, probably the most commonly measured variable is that of symptomatology or psychological/mental health status. After all, disturbance or disruption in this dimension is probably the most common reason why people seek behavioral health care services in the first place. However, there are other reasons for seeking help, including difficulties in coping with various types of life transitions (e.g., a new job, recent marriage or divorce, other changes in the work or home environment), inability to deal with the behavior of others (e.g., spouse, children), general dissatisfaction with life, or perhaps other less common reasons. Additional assessment of related variables therefore may be necessary, or may even take precedence over the assessment of symptoms or other mental health indicators.

In the vast majority of the cases seen for behavioral health care services, the assessment of the patient's overall level of psychological distress or disturbance will yield the most singularly useful information, regardless of whether it is used for outcomes measurement, outcomes monitoring, outcomes management, or to meet the requirements of third-party

payers for authorization of additional benefits. Indices such as the Positive Symptom Total (PST) or Global Severity Index (GSI) that are part of the SA-45 or BSI can provide this type of information efficiently and economically.

For some patients, measures of one or more specific psychological disorders or symptom clusters are at least as important if not more important than overall symptom or mental health status. Here, if interest is in only one disorder or symptom cluster (e.g., depression), one may choose to measure only that particular set of symptoms using an instrument designed specifically for that purpose (e.g., use of the BDI with depressed patients). For those interested in assessing the outcomes of treatment relative to multiple psychological dimensions, the administration of more than one disorder-specific instrument or a single, multidimensional instrument which assesses all or most of the dimensions of interest would be required. Again, instruments such as the SA-45 or the BSI can provide a quick, broad assessment of multiple symptom domains. Although much lengthier, other multiscale instruments, such as the MMPI-2 or the PAI, permit a more detailed assessment of several disorders or symptom domains using one inventory.

In many cases, the assessment of mental health status is adequate for outcomes assessment purposes. There are other instances in which changes in psychological distress or disturbance either provide only a partial indication of the degree to which therapeutic intervention has been successful, are not of interest to the patient or a third-party payer, are unrelated to the reason why the patient sought services in the first place, or are otherwise inadequate or unacceptable as measures of improvement in the patient's condition. One may find that for some patients, improved functioning on the job, at school, or with family or friends is much more relevant and important than symptom reduction. For other patients, improvement in their quality of life or feeling of well-being is more meaningful.

It is not always a simple matter to determine exactly what should be measured. However, careful consideration of the following questions should greatly facilitate the decision.

(i) *Why did the patient seek services?* People pursue treatment for many reasons. The patient's stated reason for seeking therapeutic assistance may be the first clue in determining what is important to measure.

(ii) *What did the patient hope to gain from treatment?* The patient's stated goals for the treatment he or she is about to receive may be a primary consideration in the selection of outcomes to be assessed.

(iii) *What are the patient's criteria for the successful completion of the current therapeutic episode?* The patient's goals for treatment may provide only a broad target for the therapeutic intervention. Having the patient identify exactly what will have to happen to consider treatment successful and no longer needed will help in specifying the most important constructs and/or behaviors to assess.

(iv) *What are the clinician's criteria for the successful completion of the current therapeutic episode?* What the patient identifies as being important to accomplish during treatment may reflect a lack of insight into his or her problems, or it might not otherwise concur with what the clinician's experience would indicate.

(v) *What are the criteria of significant third parties for the successful completion of the current therapeutic episode?* From a strict treatment perspective, this should be given the least amount of consideration. From a more realistic perspective, one cannot overlook the expectations and limitations that one or more third parties have for the treatment that is rendered. The expectations and limitations set by the patient's behavioral health care plan, the guidelines of the organization in which the clinician practices, and possibly other external forces may significantly play into the decision about when to terminate treatment.

(vi) *What, if any, are the outcomes initiatives within the provider organization?* One cannot ignore any outcomes programs that have been initiated by the organization in which the therapeutic services are delivered. Regardless of the problems and goals of the individual patient, organization-wide studies of effectiveness may dictate the gathering of specific types of outcomes data from patients who have received services.

Note that the selection of the variables to be assessed may address more than one of the above issues. Ideally, this is what should happen. However, one needs to take care that the gathering of outcomes data does not become too burdensome. As a general rule, the more outcomes data one attempts to gather from a given patient or collateral, the less likely one is to obtain any data at all. The key is to identify the point at which the amount of data that can be obtained from a patient and/or collaterals, and the ease with which it can be gathered, is optimized.

4.18.8.4.3 How to measure

Once the decision concerning what to measure has been made, one must then decide how this should be measured. In many cases, the most important data will be that obtained directly from the patient through the use of self-report instruments. Underlying this assertion are the assumptions that valid and reliable instrumentation, appropriate to the needs of the patient, is available to the clinician; the patient can read at the level required by the instruments; and the patient is motivated to respond honestly to the questions asked. If this is not the case, other options are available.

Other types of data-gathering tools may be substituted for self-report measures. Rating scales completed by the clinician or other members of the treatment staff may provide information that is as useful as that elicited directly from the patient. In those cases in which the patient is severely disturbed, unable to give valid and reliable answers (e.g., younger children), unable to read, or is an otherwise inappropriate candidate for a self-report measure, clinical rating scales can substitute as useful means of gathering data. Related to these instruments are parent-completed inventories for child and adolescent patients. These are particularly useful in obtaining information about the child or teen's behavior that might not otherwise be known.

Collateral rating instruments can also be used to gather information in addition to that obtained from self-report measures. When used in this manner, these instruments provide a mechanism by which the clinician and other treatment staff can contribute data to the outcomes assessment endeavor. This not only results in the clinician or provider organization having more information upon which to evaluate the outcomes of therapeutic intervention, it also gives the clinician an opportunity to ensure that the perspective of the treatment provider is considered in the evaluation of the effects of the treatment given.

Another potential source of outcomes information is administrative data. In many of the larger provider organizations, this information is easily retrieved through their management information systems (MISs). Data related to the patient's diagnosis, dose and regimen of medication, physical findings, course of treatment, and other types of data typically stored in these systems can be useful to those evaluating the outcomes of therapeutic intervention.

4.18.8.4.4 When to measure

There are no hard and fast rules, guidelines, or accepted conventions related to when outcomes should be assessed. The common practice is to assess the patient at least at treatment initiation and treatment termination/discharge. Obviously, at the time of treatment initiation, the clinician should obtain a baseline measure of whatever variables will be measured at the termination. At the minimum, this allows for "outcomes measurement" as described above. As has been discussed, additional assessment of the patient on the variables of interest can take place at other points in time, that is, at other times during the course of treatment and upon post-discharge follow up.

Many would argue that postdischarge/post-termination follow-up assessment provides the best or most important indication of the outcomes of therapeutic intervention. Two types of comparisons may be made on follow-up. The first is a comparison of the patient's status on the variables of interest at the time of treatment initiation, or at the time of discharge or termination, to that of the patient at some point after treatment has ended. Either way, this follow-up data will provide an indication of the more lasting effects of the intervention. Generally, the variables of interest for this type of comparison include such things as symptom presence and intensity, feeling of well-being, frequency of substance use, and social and role functioning.

The second type of post-treatment investigation involves looking at the frequency at which some aspect(s) of the patient's life circumstances, behavior or functioning occurred during a given period prior to treatment, compared to that which occurred during an equivalent period of time immediately preceding the post-discharge assessment. This approach is commonly used in determining the cost-offset benefits of treatment. For example, the number of times a patient has been seen in an emergency room for psychiatric problems during the three-month period preceding the initiation of outpatient treatment can be compared to the number of emergency room visits during the three-month period preceding the postdischarge follow-up assessment. Not only can this provide an indication of the degree to which treatment has helped the patient deal with his problems, it can also demonstrate how much medical expenses have been reduced through the patient's decreased use of costly emergency room services.

In general, post-discharge outcomes assessment probably should take place no sooner than a month after treatment has ended. When feasible, one probably should wait three to six months to assess the variables. This should provide a more valid indication of the lasting effects of treatment. Assessments being conducted to determine the frequency with which some behavior or event occurs (as may be needed to determine cost-offset benefits) can be accomplished no sooner than the reference time interval used in the baseline assessment. Thus,

suppose that the patient reports 10 emergency room visits during the three-month period prior to treatment. If one wants to know if the patient's emergency room visits have decreased after treatment, the assessment cannot take place any earlier than three months after treatment termination.

4.18.8.4.5 How to analyze outcomes data

There are two general approaches to the analysis of treatment outcomes data. The first is to determine whether changes in patient scores on outcomes measures are statistically significant. The other is to establish whether these changes are clinically significant. Use of standard tests of statistical significance is important in the analysis of group or population change data. Clinical significance is more relevant to change in the individual patient. As this chapter is focused on the individual patient, this section will center on matters related to determining clinically significant change as the result of treatment.

The issue of clinical significance has received a great deal of attention in psychotherapy research during the past several years. This is at least partially owing to the work of Jacobson and his colleagues (Jacobson, Follette, & Revenstorf, 1984, 1986; Jacobson & Truax, 1991) and others (e.g., Christensen & Mendoza, 1986; Speer, 1992; Wampold & Jenson, 1986). Their work came at a time when researchers began to recognize that traditional statistical comparisons do not reveal a great deal about the efficacy of therapy. In discussing the topic, Jacobson and Truax broadly define the clinical significance of a treatment as

> its ability to meet standards of efficacy set by consumers, clinicians, and researchers. While there is little consensus in the field regarding what these standards should be, various criteria have been suggested: a high percentage of clients improving . . .; a level of change that is recognizable by peers and significant others . . .; an elimination of the presenting problem . . .; normative levels of functioning at the end of therapy . . . ; high end-state functioning at the end of therapy . . .; or changes that significantly reduce one's risk for various health problems. (p. 12)

From their perspective, Jacobson and his colleagues (Jacobson, Follette, & Revenstorf, 1984; Jacobson & Truax, 1991) felt that clinically significant change could be conceptualized in one of three ways. Thus, for clinically significant change to have occurred, the measured level of functioning following the therapeutic episode would either:

(i) fall outside the range of the dysfunctional population by at least two standard deviations away from the mean of that population, in the direction of functionality;

(ii) fall within two standard deviations of the mean for the normal or functional population; or

(iii) be closer to the mean of the functional population than to that of the dysfunctional population.

Jacobson and Truax viewed the third option as being the least arbitrary and provided different recommendations for determining cutoffs for clinically significant change, depending upon the availability of normative data. Lambert (1994) demonstrated how the third option could be modified to allow for the inclusion of more than one categorization of dysfunction (e.g., mild, moderate, severe). This assumes, of course, that the necessary normative data needed to separate the gradations of dysfunction are available.

At the same time, these same investigators noted the importance of considering the change in the measured variables of interest from pre- to post-treatment, in addition to the patient's functional status at the end of therapy. To this end, Jacobson et al. (1984) proposed the concomitant use of a reliable change (RC) index to determine whether change is clinically significant. This index, modified on the recommendation of Christensen and Mendoza (1986), is nothing more than the pretest score minus the posttest score divided by the standard error of the difference of the two scores, expressed as:

$$RC = (x_2 - x_1)/S_{diff}$$

where x_1 is the pretest score, x_2 is the post-test score, and S_{diff} is the standard error of the difference. The standard error of the difference is computed as:

$$S_{diff} = \sqrt{2\,(SEM)^2}$$

where SEM is the standard error of measurement for a functional group (e.g., normals, nonpatients) on the instrument. If the RC index is greater than 1.96, the change in scores is not likely to be due to chance ($p < 0.05$), but rather to reflect real change.

Speer (1992) recommended a different approach when regression to the mean has been demonstrated to contribute to the improvement in scores from pre- to post-test. The alternate approach, based on the combined work of Nunnally (1967) and Edwards, Yarvis, Mueller, Zingale, and Wagman (1978), involves developing a confidence interval of ± 2 SEMs around the estimated true pretest score. A post-test score falling outside of this confidence interval is considered significantly different from the initial pretest score at $p < 0.05$. Using this approach, more change is needed to show clinically significant improvement than to show clinically significant deterioration. Note that the criterion for determining whether regression to the mean is occurring is met when a negative correlation is found to exist between the pretreatment score and amount of change that has taken place. This implies the evaluation of group data, and for this reason this empirical criterion may not be of use for the individual patient unless the latter is a member of a sample for which test results are available.

Lambert (1994) proposes a modified recommendation for the dual criteria for clinically significant change (that is, RC greater than 1.96 and movement of the patient's score from the dysfunctional group's distribution to the functional group's distribution) such that movement from one degree of dysfunction to a lesser degree would also meet one of the two criteria for clinically significant change. In an example, Lambert illustrated that normative data for the Global Severity Index (GSI) found in the SCL-90-R literature can be used to empirically define four levels of symptom intensity: asymptomatic, mildly symptomatic, moderately symptomatic, and severely symptomatic. Assuming an RC of 1.96 or greater, clinically significant change can be said to have occurred if a patient's GSI score moves from severely to moderately or mildly symptomatic, or to asymptomatic; from moderately to mildly symptomatic, or to asymptomatic; or from mildly symptomatic to asymptomatic. Although this criterion is less stringent than having to move from being symptomatic (regardless of the severity) to asymptomatic, it still provides information that is quite useful for clinical decision making.

4.18.9 FUTURE DIRECTIONS

The ways in which clinical psychologists have conducted the types of psychological assessment described in this chapter have undergone dramatic changes during the 1990s. This should come as no surprise to anyone who spends a few minutes a day skimming the newspaper or watching television. The health care revolution started gaining momentum at the beginning of the 1990s and has not since slowed down. And there are no indications that it will subside in the foreseeable future. There was no real reason to think that behavioral health care would be spared from being a target of the revolution, and there is no good reason why it should have been spared. The behavioral health care industry certainly had contributed its share of waste, inefficiency, and lack of accountability to the problems that led to the revolution. Now, like other areas of health care, it is forced to "clean up its act." Some consumers of mental health or chemical dependency services have benefited from the revolution, others have not. In any case, the way in which health care is delivered and financed has changed, and clinical psychologists and other behavioral health care professionals must adapt to survive in the market.

Some of those involved in the delivery of psychological assessment services may wonder (with some fear and trepidation) where the revolution is leading the behavioral health care industry and, in particular, how their ability to practice will be affected. At the same time, others are eagerly awaiting the inevitable advances in technology and other resources that come with the passage of time. What will occur is open to speculation. However, close observation of the practice of psychological assessment and the various industries that support it (particularly the forms of therapeutic assessment described in this chapter) has led this author to arrive at some predictions as to where the field of therapeutic psychological assessment is headed and the implications these have for clinicians, provider organizations, and patients. What follows in this section are the most important of these predictions. Also included are what this author feels are the needs that must be met if psychological assessment is to continue to be a valued contributor to the delivery of efficient, cost-effective behavioral health care.

4.18.9.1 What the Industry Is Moving Away From?

One way of discussing what the field is moving toward is to first talk about what it is moving away from. In the case of therapeutic psychological assessment, two trends are becoming quite clear. First, starting at the beginning of this last decade of the twentieth century, the use of (and reimbursement for) psychological assessment has gradually been curtailed. This particularly has been the case with regard to indiscriminate assessment involving the administration of lengthy and expensive batteries of psychological tests. Payers began to demand evidence that the knowledge gained from the administration of these instruments contributed to the delivery of cost-effective, efficient care to mental health and substance abuse patients. There are no indications that this trend will stop.

Second, assessment has begun to move away from the use of lengthy, multidimensional objective instruments (e.g., the MMPI) or time-consuming projective techniques (e.g., Rorschach) that previously represented the standard of practice. When assessment is authorized now, it usually involves the use of inexpensive yet well-validated, problem-oriented instruments. This reflects modern behavioral health care's time-limited, problem-oriented approach to treatment. The clinician can no longer afford to spend a great deal of time in assessment activities when the patient has only a limited number of payer-authorized sessions with him or her. Thus, both now and in the foreseeable future, brief instruments will be used for problem identification or clarification, progress monitoring, and/or outcomes assessment.

4.18.9.2 Trends in Instrumentation

The move toward the use of brief, problem-oriented instruments for therapeutic psychological assessment purposes has just been identified. Another trend in the selection of instrumentation is the increasing use of public domain tests, questionnaires, rating scales, and other types of measurement tools. Previously, these free-use instruments were not developed with the rigor that is usually applied in the development of psychometrically sound instruments by commercial test publishers. Consequently, they typically lacked the validity and reliability data that are necessary to judge their psychometric integrity.

Recently, however, there has been a significant improvement in the quality and documentation of the public domain and other "for-free" tests that are available for therapeutic psychological assessment. Instruments such as the SF-36/SF-12 and HSQ/HSQ-12 health measures are good examples of such tools. These and instruments such as the Behavior and Symptom Identification Scale (BASIS-32; Eisen, Grob, & Klein, 1986) and the Outcome Questionnaire (OQ-45.1; Lambert, Lunnen, Umphress, Hansen, & Burlingame, 1994) have undergone psychometric scrutiny and have gained widespread acceptance. Although copyrighted, they may be used for a nominal one-time or annual licensing fee; thus, they generally are treated much like public domain assessment tools. One can expect that other quality, useful instruments will be made available for use at little or no cost in the future.

As for the types of instrumentation that will be needed and developed, one can probably expect some changes. Accompanying the increasing focus on outcomes assessment is a recognition by payers and patients that changes in several areas of functioning are at least as important as changes in level of symptom severity when evaluating the effectiveness of the treatment. For example, employers are interested in the patient's ability to resume the functions of his or her job, while family members may be concerned with the patient's ability to resume their role as spouse or parent. Increasingly, measurement of the patient's functioning in areas other than psychological/mental status has come to be included as part of behavioral health care outcomes systems. Probably the most visible indication of this is the incorporation of the SF-36 or HSQ into various behavioral health care studies, and the fact that two major psychological test publishers offer HSQ products in their catalogs of clinical products. One will likely see other public domain and commercially available nonsymptom-oriented instruments, especially those emphasizing social and occupational role functioning, in increasing numbers over the next several years.

Other types of instrumentation will also become prominent. These will include measures of variables that support the outcomes and other therapeutic assessment initiatives undertaken by provider organizations. What one organization or provider feels is important, or what it is told is important for reimbursement or other purposes, will dictate what is measured. Instrumentation will also include measures that will be useful for the prediction of outcomes for individuals seeking psychotherapeutic services from those organizations.

4.18.9.3 Trends in Data Use and Storage

There are two areas of application in which the valuable data obtained from therapeutic psychological assessment have heretofore been overlooked or underutilized. Indications are that this will change for both in the future. One area for which assessment data has potential application is that of clinical decision-making. This of course pertains only to the use of outcomes assessment data. Generally, data gathered solely for the purpose of outcomes assessment is used for just that: the assessment of the results of treatment. This is particularly the case in large, formal outcomes management programs. As has been discussed earlier in this chapter, data gathered at the beginning of treatment can be used immediately for treatment planning purposes while also serving as baseline data that can be compared to discharge data later on.

The other potential area of data application is in the development of local, regional, and national databases of therapeutic assessment data. Patient data gathered by various providers, organizations or programs within organizations, at one or more points during the therapeutic episode, can be pooled and used for various purposes. These databases can then serve as the bases for two highly beneficial (and probably profitable) endeavors. The first is the generation of sets of normative data for various populations delineated along any number of parameters. Norms for any number of instruments or health care variables could be generated "on demand" and continuously updated to reflect trends in behavioral health care. This author is aware of one large, national behavioral health care system where such a database already exists. He also is aware of efforts at establishing cross-organizational databases of this kind.

The second benefit afforded by the information contained in these databases is that of predictive modeling. For example, the behavioral health care organization mentioned above has taken advantage of the organizational data available to it to investigate the relationships between a number of treatment, demographic and other variables and the outcomes of treatment. Subjecting the large data sets available to it to sophisticated statistical analyses has allowed this organization to determine those types of patients requiring special care or attention in order to achieve desired outcomes at the time of treatment termination. Predictive modeling can also be used for identifying variables related to other aspects of patient care, such as patient satisfaction with the care received. The possibilities for the use of data in this manner are enormous.

4.18.9.4 Trends in the Application of Technology

Clinical psychologists have not been shy when it has come to taking advantage of the technological advances that have been achieved since the late 1970s. This is no more evident than in the extent to which the personal computer and the vast array of psychological assessment software have been incorporated into their delivery of clinical services. Automated administration, scoring, and interpretation and reporting of the results of nearly all major objective tests are currently available to the clinician through PC-based software. In addition, the availability of affordable desktop optical scanners allows the clinician to maintain the portability of the assessment instruments while retaining the scoring and interpreting power of the computer for processing the test data.

To speculate on how technology will be advanced in the service of therapeutic psychological assessment in the future is a risky businesss. As has been witnessed since the late 1970s, much can happen quickly. There are, however, three areas of technologic or technology-dependent advances on the horizon to which clinical psychologists should have access in the not too distant future. The first is the availability of online administration, scoring, and interpretation and reporting of tests via the Internet. In fact, an Internet version of the SA-45 is being beta tested at the time of writing (1997). To this author's knowledge, this represents the first use of the Internet for psychological assessment purposes. It is anticipated that the Internet version of the SA-45 will be commercially available in the very near future, and it will be quickly followed by the availability of Internet versions of other assessment instruments.

The second advance is actually a technology that has been around for a while but has undergone improvements, that is, the fax-back technology that is being used for scoring and reporting of objective, paper-and-pencil tests. Essentially, the fax machine replaces the optical scanner as a means of data entry. The electronic data is entered directly into the test publisher's computer for processing and report generation. However, instead of generating a hard-copy report of results at the processing site, the report is transmitted in electronic form and sent back to the clinician's fax machine within minutes of processing. At that point, the report is printed out just like any other fax transmission. Currently, this technology is being used on a somewhat limited basis. This is partially owing to the degree to which test publishers are making this form of automated scoring and reporting available to their customers. However, in the relatively near future, one should see more tests being offered to clinicians in this manner, particularly as the technology continues to improve.

The third area of technologic advancement has more to do with the application of technology than the development of new technology. L. E. Beutler and O. B. Williams (personal communication, January 15, 1996) have taken Beutler and Clarkin's (1990) Systematic Treatment Selection (STS) model of prescriptive treatment assignment and have developed specifications for software for automating the matching of treatments, therapists, and patients. The capability of subsequently tracking patients during the course of treatment is also included in these specifications. Originally entitled *STS for Windows*, this software is now under development through a behavioral

health care publishing and consulting company. Driving the STS system is patient assessment data related to six variables: subjective distress, functional severity, problem complexity, potential for therapeutic resistance, coping style, and social support. Each of these variables may be assessed through either commercially available self-report instruments or clinician rating scales developed specifically for use with STS.

When fully developed, the STS should serve as the standard for in-office treatment–patient–therapist matching and patient-tracking software. According to the developers (L. E. Beutler & O. B. Williams, personal communication, January 15, 1996), the software will include numerous features, the most important of which will be: a comprehensive treatment planning report with up-to-date references to relevant research articles and treatment manuals; automatic entry of each patient's data into a growing database that is used for treatment planning and prediction; the ability to predict the amount of symptom reduction from a specific course of therapy; a report profiling the patient's symptom status over time; a report indicating individual clinician's ability to treat specific types of symptomatology; and the ability to incorporate case notes into the patient's electronic file. The major benefits of the system include the ability to: use different assessment means (self-report or clinician rating) to obtain the information needed to drive the system; develop treatment recommendations based on information that is optimal for the patient; easily monitor patient progress on a glide path developed from the treatment of similar patients and adjust the therapy plan (if necessary) on a timely basis; and determine a clinician's therapeutic strengths and weaknesses, thus permitting the most effective patient–therapist match.

All in all, when fully developed, the STS software will combine the knowledge and expertise of a leader in the field of psychotherapeutic research with state-of-the-art technology, thus yielding a powerful decision-making behavioral health care tool. One can be assured that similar products are likely to follow once the benefits of the STS software become widely known.

4.18.10 SUMMARY

The health care revolution has brought mixed blessings to those in the behavioral health care professions. It has resulted in limitations for reimbursement for services rendered and has forced many to change the way they practice their profession. At the same time, it has led to revelations about the cost savings benefits that can accrue from the treatment of mental health and substance use disorders. This has been the bright spot in an otherwise bleak picture for some behavioral health care professionals. For clinical psychologists, the picture appears to be somewhat different. They now have additional opportunities to contribute to the positive aspects of the revolution and to gain from the "new order" it has imposed. By virtue of their training in psychological assessment and through the application of appropriate instrumentation, they are uniquely qualified to support or otherwise facilitate multiple aspects of the therapeutic process. It is the clinical psychologist's contributions to aspects of "therapeutic psychological assessment" that this chapter has sought to identify and address in some detail.

Earlier, this author identified some of the types of psychological assessment instruments that are commonly used in the service of therapeutic endeavors. These included both brief and lengthy (multidimensional) symptom measures, as well as measures of general health status, quality of life, and patient satisfaction with the services received. Also identified were different sets of general criteria that can be applied when selecting instruments for use in therapeutic settings. The main intent of this chapter, however, was to present a detailed discussion of the various therapeutic uses of psychological assessment.

Generally, psychological assessment can assist the clinician in three important clinical activities: clinical decision-making, treatment itself (when used as a specific therapeutic technique), and treatment outcomes assessment. Regarding the first of these activities, three important clinical decision-making functions can be facilitated by psychological assessment: screening, treatment planning, and treatment monitoring. The first of these can be served by the use of "down and dirty" instruments to identify, within a high degree of certainty, the likelihood of the presence (or absence) of a particular condition or characteristic. Here, the diagnostic efficiency of the instrument used (as indicated by the PPP and NPP) is of great importance. Through their ability to identify and clarify problems as well as other important treatment-relevant patient characteristics, psychological assessment instruments can also be of great assistance in planning treatment. In addition, treatment monitoring, or the regular determination of the patient's progress throughout the course of treatment, can be served well by the application of psychological assessment instruments.

Secondly, assessment may be used as part of a therapeutic technique. In what Finn terms "therapeutic assessment," situations in which patients are evaluated via psychological testing are used as opportunities for the process itself to serve as a form of therapeutic intervention. This is accomplished through involving the patient as an active participant in the assessment process, not just as the object of the assessment.

Thirdly, psychological assessment can be employed as the primary mechanism by which the outcomes or results of treatment can be measured. However, the use of assessment for this purpose is not a cut-and-dried matter. As discussed, there are issues, pertaining to what to measure, how to measure, and when to measure, that require considerable thought prior to undertaking any standard (or even nonstandard) plan to assess outcomes. Guidelines for resolving these issues are presented, as is information pertaining to how to determine whether the measured outcomes of treatment are indeed "significant."

In the final section of the chapter, this author shares some thoughts about where psychological assessment is probably heading in the future. No radical revelations are presented since no signs really point in that direction. What is foreseen is the appearance of more quality assessment instruments that will remain in the public domain, and greater application of communications technology, fax and the Internet, in particular, as assessment delivery, scoring and reporting mechanisms. Also predicted is the application of tomorrow's computer technology to available assessment data for optimized treatment–patient–therapist matching. The innovative proposals of Beutler and Williams in this regard seem to represent the state-of-the-art thinking at this time.

There is no doubt that the practice of psychological assessment has been dealt a blow within recent years. However, as this chapter hopefully has shown, clinical psychologists have the skills to take this powerful tool, apply it in ways that will benefit those suffering from mental health and substance abuse problems, and demonstrate its benefits and their skills to patients and payers. Whether they will be successful in this demonstration will be determined in the near future. In the meantime, advances will continue to be made that will facilitate their work and improve its quality.

4.18.11 REFERENCES

American Psychological Association (1992). *Ethical principles of psychologists*. Washington, DC: Author.

American Psychological Association (1996). *The costs of failing to provide appropriate mental health care*. Washington, DC: Author.

Andrews, G., Peters, L., & Teesson, M. (1994). *The measurement of consumer outcomes in mental health*. Canberra, Australia: Australian Government Publishing Service.

Appelbaum, S. A. (1990). The relationship between assessment and psychotherapy. *Journal of Personality Assessment, 54*, 791–801.

Attkisson, C. C., & Zwick, R. (1982). The Client Satisfaction Questionnaire: Psychometric properties and correlations with service utilization and psychotherapy outcome. *Evaluation and Program Planning, 6*, 233–237.

Baldessarini, R. J., Finkelstein, S., & Arana, G. W. (1983). The predictive power of diagnostic tests and the effect of prevalence of illness. *Archives of General Psychiatry, 40*, 569–573.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford Press.

Beutler, L. E., & Clarkin, J. (1990). *Systematic treatment selection: Toward targeted therapeutic interventions*. New York: Brunner/Mazel.

Beutler, L. E., Wakefield, P., & Williams, R. E. (1994). Use of psychological tests/instruments for treatment planning. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (pp. 55–74). Hillsdale, NJ: Erlbaum.

Beutler, L. E., & Williams, O. B. (1995). Computer applications for the selection of optimal psychosocial therapeutic interventions. *Behavioral health care Tomorrow, 4*, 66–68

Burlingame, G. M., Lambert, M. J., Reisinger, C. W., Neff, W. M., & Mosier, J. (1995). Pragmatics of tracking mental health outcomes in a managed care setting. *Journal of Mental Health Administration, 22*, 226–236.

Butcher, J. N. (1990). *The MMPI-2 in psychological treatment*. New York: Oxford University Press.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A. M., & Kaemmer, B. (1989). *MMPI-2: Manual for administration and scoring*. Minneapolis, MN: University of Minnesota Press.

Butcher, J. N., Graham, J. R., Williams, C. L., & Ben-Porath, Y. (1989). *Development and use of the MMPI-2 content scales*. Minneapolis, MN: University of Minnesota Press.

Cagney, T., & Woods, D. R. (1994). Why focus on outcomes data? *Behavioral health care Tomorrow, 3*, 65–67.

Center for Disease Control and Prevention (1994, May 27). Quality of life as a new public health measure: Behavioral risk factor surveillance system. *Morbidity and Mortality Weekly Report, 43*, 375–380.

Christensen, L., & Mendoza, J. L. (1986). A method of assessing change in a single subject: An alteration of the RC index [Letter to the editor]. *Behavior Therapy, 17*, 305–308.

Ciarlo, J. A., Brown, T. R., Edwards, D. W., Kiresuk, T. J., & Newman, F. L. (1986). *Assessing mental health treatment outcomes measurement techniques*. DHHS Pub. No. (ADM)86–1301. Washington, DC: US Government Printing Office.

Commission on Chronic Illness (1987). *Chronic illness in the United States: 1*. Cambridge, MA: Commonwealth Fund, Harvard University Press.

Derogatis, L. R. (1983). *SCL-90-R: Administration, scoring and procedures manual-II*. Baltimore: Clinical Psychometric Research.

Derogatis, L. R. (1992). *BSI: Administration, scoring and procedures manual-II*. Baltimore: Clinical Psychometric Research.

Derogatis, L. R., & DellaPietra, L. (1994). Psychological tests in screening for psychiatric disorder. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (pp. 22–54). Hillsdale, NJ: Erlbaum.

F.P.C.E. BIBLIOTECA

Derogatis, L. R., Lipman, R. S., & Covi, L. (1973). SCL-90: An outpatient psychiatric rating scale—preliminary report. *Psychopharmacology Bulletin, 9,* 13–27.

Dickey, B., & Wagenaar, H. (1996). Evaluating health status. In L. I. Sederer & B. Dickey (Eds.), *Outcomes assessment in clinical practice* (pp. 55–60). Baltimore: Williams & Wilkins.

Donabedian, A. (1985). *Explorations in quality assessment and monitoring: Vol. III. The methods and findings of quality assessment monitoring: An illlustrative analysis.* Ann Arbor, MI: Health Administration Press.

Dorwart, R. A. (1996). Outcomes management strategies in mental health: Applications and implications for clinical practice. In L. I. Sederer & B. Dickey (Eds.), *Outcomes assessment in clinical practice* (pp. 45–54). Baltimore: Williams & Wilkins.

Dowd, E. T., Milne, C. R., & Wise, S. L. (1991). The therapeutic Reactance Scale: A measure of psychological reactance. *Journal of Counseling and Development, 69,* 541–545.

Edwards, D. W., Yarvis, R. M., Mueller, D. P., Zingale, H. C., & Wagman, W. J. (1978). Test-taking and the stability of adjustment scales: Can we assess patient deterioration? *Evaluation Quarterly, 2,* 275–292.

Eisen, S. V., Grob, M. C., & Klein, A. A. (1986). BASIS: The development of a self-report measure for psychiatric inpatient evaluation. *The Psychiatric Hospital, 17,* 165–171.

Elwood, R. W. (1993). Psychological tests and clinical discrimination: Beginning to address the base rate problem. *Clinical Psychology Review, 13,* 409–419.

Ficken, J. (1995). New directions for psychological testing. *Behavioral Health Management, 20,* 12–14.

Finn, S. E. (1996a). *Manual for using the MMPI-2 as a therapeutic intervention.* Minneapolis, MN: University of Minnesota Press.

Finn, S. E. (1996b). Assessment feedback integrating MMPI-2 and Rorschach findings. *Journal of Personality Assessment, 67,* 543–557.

Finn, S. E., & Butcher, J. N. (1991). Clinical objective personality assessment. In M. Hersen, A. E. Kazdin, & A. S. Bellack (Eds.), *The clinical psychology handbook* (2nd ed., pp. 362–373). New York: Pergamon.

Finn, S. E., & Martin, H. (in press). Therapeutic assessment with the MMPI-2 in managed health care. In J. N. Butcher (Ed.), *Objective personality assessment in managed health care: A practitioner's guide.* Minneapolis, MN: University of Minnesota Press.

Finn, S. E., & Tonsager, M. E. (1992). Therapeutic effects of providing MMPI-2 test feedback to college students awaiting therapy. *Psychological Assessment, 4,* 278–287.

Friedman, R., Sobel, D., Myers, P., Caudill, M., & Benson, H. (1995). Behavioral medicine, clinical health psychology, and cost offset. *Health Psychology, 14,* 509–518.

Gough, H. G., McClosky, H., & Meehl, P. E. (1951). A personality scale for dominance. *Journal of Abnormal and Social Psychology, 46,* 360–366.

Gough, H. G., McClosky, H., & Meehl, P. E. (1952). A personality scale for social responsibility. *Journal of Abnormal and Social Psychology, 47,* 73–80.

Greene, R. L. (1991). *The MMPI-2/MMPI: An interpretive manual.* Boston: Allyn & Bacon.

Greene, R. L., & Clopton, J. R. (1994). Minnesota Multiphasic Personality Inventory—2. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (pp. 137–159). Hillsdale, NJ: Erlbaum.

Greenfield, T. K., & Attkisson, C. C. (1989). Progress toward a multifactorial service satisfaction scale for evaluating primary care and mental health services. *Evaluation and Program Planning, 12,* 271–278.

Hathaway, S. R., & McKinley, J. C. (1951). *MMPI manual.* New York: The Psychological Corporation.

Health Outcomes Institute (1993). *Health Status Questionnaire 2.0 manual.* Bloomington, MN: Author.

Hsiao, J. K., Bartko, J. J., & Potter, W. Z. (1989). Diagnosing diagnoses: Receiver operating characteristic methods and psychiatry. *Archives of General Psychiatry, 46,* 664–667.

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15,* 336–352.

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1986). Toward a standard definition of clinically significant change [Letter to the editor]. *Behavior Therapy, 17,* 309–311.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12–19.

Jahoda, M. (1958). *Current concepts of mental health.* New York: Basic Books.

Lambert, M. J. (1994). Use of psychological tests for outcome assessment. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (pp. 75–97). Hillsdale, NJ: Erlbaum.

Lambert, M. J., Lunnen, K., Umphress, V., Hansen, N. B., & Burlingame, G. M. (1994). *Administration and scoring manual for the Outcome Questionnaire (OQ-45.1).* Salt Lake City, UT: IHC Center for Behavioral health care Efficacy.

Larsen, D. L., Attkisson, C. C., Hargreaves, W. A., & Nguyen, T. D. (1979). Assessment of client/patient satisfaction: Development of a general scale. *Evaluation and Program Planning, 2,* 197–207.

LeVois, M., Nguyen, T. D., & Attkisson, C. C. (1981). Artifact in client satisfaction assessment: Experience in community mental health settings. *Evaluation and Program Planning, 4,* 139–150.

Maruish, M. (1990, Fall). Psychological assessment: What will its role be in the future? *Assessment Applications,* 7–8.

Maruish, M. E. (1994). Introduction. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (pp. 3–21). Hillsdale, NJ: Erlbaum.

Megargee, E. I., & Spielberger, C. D. (1992). Reflections on fifty years of personality assessment and future directions for the field. In E. I. Megargee & C. D. Spielberger (Eds.), *Personality assessment in America* (pp. 170–190). Hillsdale, NJ: Erlbaum.

Mental Health Weekly. (1996a, April 8). Future targets behavioral health field's quest for survival, 1–2.

Mental Health Weekly. (1996b, April 8). Leaders predict integration of MH primary care by 2000, 1–6.

Mental Health Weekly. (1996c, April 29). Anxiety disorders screening day occurs this week, 7.

Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine, 8,* 283–298.

Migdail, K. J., Youngs, M. T., & Bengen-Seltzer, B. (Eds.) (1995). *The 1995 behavioral outcomes & guidelines sourcebook.* New York: Faulkner & Gray.

Millon, T. (1994). *MCMI-III manual.* Minneapolis, MN: National Computer Systems.

Moreland, K. L. (1996). How psychological testing can reinstate its value in an era of cost containment. *Behavioral health care Tomorrow, 5,* 59–61.

Morey, L. C. (1991). *The Personality Assessment Inventory professional manual.* Odessa, FL: Psychological Assessment Resources.

Morey, L. C., & Henry, W. (1994). Personality Assessment Inventory. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (pp. 185–216). Hillsdale, NJ: Erlbaum.

Newman, F. L. (1994). Selection of design and statistical procedures for progress and outcome assessment. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (pp. 111–134). Hillsdale, NJ: Erlbaum.

Newman, F. L., & Ciarlo, J. A. (1994). Criteria for selecting psychological instruments for treatment outcome assessment. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (pp. 98–110). Hillsdale, NJ: Erlbaum.

Nguyen, T. D., Attkisson, C. C., & Stegner, B. L. (1983). Assessment of patient satisfaction: Development and refinement of a service evaluation questionnaire. *Evaluation and Program Planning, 6,* 299–313.

Nunnally, J. C. (1967). *Psychometric theory.* New York: McGraw-Hill.

Olfson, M., & Pincus, H. A. (1994a). Outpatient psychotherapy in the United States, I: Volume, costs, and user characteristics. *American Journal of Psychiatry, 151,* 1281–1288.

Olfson, M., & Pincus, H. A. (1994b). Outpatient psychotherapy in the United States, II: Patterns of utilization. *American Journal of Psychiatry, 151,* 1289.

Oss, M. E. (1996). Managed behavioral health care: A look at the numbers. *Behavioral Health Management, 16,* 16–17.

Pallak, M. S. (1994). National outcomes management survey: Summary report. *Behavioral health care Tomorrow, 3,* 63–69.

Phelps, R. (1996, February). Preliminary practitioner survey results enhance APA's understanding of health care environment. *Practitioner Focus, 9,* 5.

Psychotherapy Finances (1995, January). Fee, practice and managed care survey. 21(1), Issue 249.

Radosevich, D., & Pruitt, M. (1996). *Twelve-item Health Status Questionnaire (HSQ-12) version 2.0 user's guide.* Bloomington, MN: Health Outcomes Institute.

Radosevich, D. M., Wetzler, H., & Wilson, S. M. (1994). *Health Status Questionnaire (HSQ) 2.0: Scoring comparisons and reference data.* Bloomington, MN: Health Outcomes Institute.

Schlosser, B. (1995). The ecology of assessment: A "patient-centric" perspective. *Behavioral health care Tomorrow, 4,* 66–68.

Sederer, L. I., Dickey, B., & Hermann, R. C. (1996). The imperative of outcomes assessment in psychiatry. In L. I. Sederer & B. Dickey (Eds.), *Outcomes assessment in clinical practice* (pp. 1–7). Baltimore: Williams & Wilkins.

Simmons, J. W., Avant, W. S., Demski, J., & Parisher, D. (1988). Determining successful pain clinic treatment through validation of cost effectiveness. *Spine, 13,* 34.

Sipkoff, M. Z. (1995, August). Behavioral health treatment reduces medical costs: Treatment of mental disorders and substance abuse problems increases productivity in the workplace. *Open Minds,* 12.

Speer, D. C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting and Clinical Psychology, 60,* 402–408.

Spielberger, C. D. (1983). *Manual of the State–Trait Anxiety Inventory: STAI (Form Y).* Palo Alto, CA: Consulting Psychologists Press.

Stewart, A. L., & Ware, J. E., Jr. (1992). *Measuring functioning and well-being.* Durham, NC: Duke University Press.

Strain, J. J., Lyons, J. S., Hammer, J. S., Fahs, M., Lebovits, A., Paddison, P. L., Snyder, S., Strauss, E., Burton, R., & Nuber, G. (1991). Cost offset from a psychiatric consultation-liaison intervention with elderly hip fracture patients. *American Journal of Psychiatry, 148,* 1044–1049.

Strategic Advantage, Inc. (1996). *Symptom Assessment-45 Questionnaire manual.* Minneapolis, MN: Author.

Substance Abuse Funding News, (1995, December 22). Brown resigns drug post. 7.

VandenBos, G. R., DeLeon, P. H., & Belar, C. D. (1993). How many practitioners are needed? It's too early to know! *Professional Psychology: Research and Practice, 22,* 441–448.

Vermillion, J., & Pfeiffer, S. (1993). Treatment outcome and continuous quality improvement: Two aspects of program evaluation. *Psychiatric Hospital, 24,* 9–14.

Wampold, B. E., & Jenson, W. R. (1986). Clinical significance revisited [Letter to the editor]. *Behavior Therapy, 17,* 302–305.

Ware, J. E., Kosinski, M., & Keller, S. D. (1995). *SF-12: How to Score the SF-12 Physical and Mental summary scales* (2nd ed.). Boston: New England Medical Center, The Health Institute.

Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-Item Short Form Health Survey (SF-36). I. Conceptual framework and item selection. *Medical Care, 30,* 473–483.

Ware, J. E., Snow, K. K., Kosinski, M., & Gandek, B. (1993). *SF-36 Health Survey manual and interpretation guide.* Boston: New England Medical Center, The Health Institute.

Werthman, M. J. (1995). A managed care approach to psychological testing. *Behavioral Health Management, 15,* 15–17.

World Health Organization (1948). Constitution. In *Basic Documents.* Geneva, Switzerland: Author.

Chapter 15

PSYCHOLOGICAL REPORT

The psychological report is the end product of assessment. It represents the clinician's efforts to integrate the assessment data into a functional whole so that the information can help the client solve problems and make decisions. Even the best tests are useless unless the data from them is explained in a manner that is relevant and clear, and meets the needs of the client and referral source. This requires clinicians to give not merely test results, but also interact with their data in a way that makes their conclusions useful in answering the referral question, making decisions, and helping to solve problems.

An evaluation can be written in several possible ways. The manner of presentation used depends on the purpose for which the report is intended as well as on the individual style and orientation of the practitioner. The format provided in this chapter is merely a suggested outline that follows common and traditional guidelines. It includes methods for elaborating on essential areas such as the referral question, behavioral observations, relevant history, impressions (interpretations), and recommendations. This format is especially appropriate for evaluations that are problem oriented and that offer specific prescriptions for change. Additional alternatives for organizing the report are to use a letter format, give only the summary and recommendations, focus on a specific problem, summarize the results test by test, write directly to the client, or provide client descriptions around a particular theory of personality. The sample evaluations vary somewhat from the suggested format, although they usually still include the essential categories of information that are discussed in this chapter.

One general style to avoid is sometimes referred to as a "shotgun" report (Tallent, 1992, 1993). This provides a wide variety of often-fragmented descriptions in the hope that some useful information can be found within. The shotgun approach is usually vague, stereotyped, and overinclusive. The recommendations for treatment are often neither specific nor practical. The most frequent reason for a shotgun report is a referral question that is too general, vague, and, therefore, poorly understood. In contrast, the "case-focused" report centers on the specific problems outlined by the referring person. It reveals unique aspects of the client and provides specific accurate descriptions, rather than portraying stereotypes that may also be overly "theory linked" or "test linked." Furthermore, the recommendations for treatment are both specific and practical. The general approach of the case-focused report is not so much *what* is to be known, but rather *why* different types of information are important for the purposes of the report.

The creation of a case-focused report involves understanding and applying several basic principles. First, the report should use action-oriented language rather than

metapsychological abstractions. This means the client's ongoing behaviors and likely personality processes should be described in relation to different situations. The use of "action-language" links the person with specific behaviors, forces reports to address specific therapeutic issues, and conveys a better understanding of the client's active role in the testing situation. Second, the recommendations in a case-focused report need to directly relate to what specifically can be done for this client in his or her particular environment. They may apply to areas such as occupational choice, psychotherapy, institutional programs, or additional evaluation. In certain types of referrals, however, especially clients self-referred for psychotherapy, an important goal may be to help them increase their level of personal insight. In these cases, a wider description of the client that includes a number of different topics might be more appropriate than the narrower, problem-solving approach. In addition, there should be a focus on that which differentiates one person from another. This means avoiding discussions of what is average about the client, and emphasizing instead what stands out and is unique to this individual. Further, there is a current trend, consistent with the case-focused approach, toward de-emphasizing diagnosis and etiology. There is, rather, an emphasis on current descriptions of the person that are tied to specific behaviors. In certain cases, especially in a medical setting, the clinician may still need to provide diagnoses in addition to behaviorally oriented descriptions. Another consideration is that a case-focused report should be written with an awareness of the point of view of the intended readers. This includes taking into consideration their level of expertise, their theoretical or professional orientation, the decisions they are facing, and the possible interpretations they are likely to make of the information.

A final point is that the quality and usefulness of a report is typically enhanced if the practitioner is knowledgeable about the area or type of issue the client is experiencing. Such knowledge helps to increase the depth of the interpretations and provides relevant information or a general "map" of the problem area that can be used to help ensure that all relevant aspects have been covered. Importantly, background knowledge on the problem area provides relevant information on a range of interventions as well as the effectiveness of these interventions. For example, knowledge regarding depression means that the practitioner is aware of its causes, variety of ways in which it is expressed, options for interventions, and when further assessment is indicated (for suicide potential). Often consulting a well-written, up-to-date chapter will provide sufficient information. In the general clinical area, useful resources are Barlow's (2001) *Clinical Handbook of Psychological Disorders* (4th ed.), R. Meyer's (1996) *The Clinician's Handbook* (3rd ed.), or Kaplan and Sadock's (2001) *Kaplan & Sadock's Pocket Handbook to Clinical Psychiatry* (3rd ed.). Persons preparing neuropsychological reports might consult Groth-Marnat's (2000a) *Neuropsychology Assessment in Clinical Practice: A Guide to Test Interpretation and Integration,* Lezak's (1995) *Neuropsychological Assessment,* or Snyder and Nussbaum's (1998) *Clinical Neuropsychology: A Pocket Handbook for Assessment.* Educational report writers might benefit from reading relevant sections in Sattler's (2001, 2002) *Assessment of Children* or Walker and Roberts' (2001) *Handbook of Clinical Child Psychology.* A particularly useful resource when doing vocational assessments is Lowman's (1991) *The Clinical Practice of Career Assessment: Interests, Abilities, and Personality.*

GENERAL GUIDELINES

Length

The typical psychological report is between five and seven single-spaced pages (Finn, Moes, & Kaplan, 2001). However, the length can vary substantially based on the purpose of the report, context, and expectations of the referral source. In medical contexts, a two-page report is not uncommon. This parallels the format of many physician reports that are a similar length. In contrast, legal contexts often require reports that are from seven to ten pages because of the greater need for documentation combined with more extensive referral questions. It is not unusual for a psychologist serving as an expert witness to not only evaluate a client, but also anticipate and defend himself or herself against rebuttals as well as comment on reports made by other mental health professionals. The more moderate (and frequent) five- to seven-page report is particularly prevalent in psychological, educational, and vocational contexts.

Style

The style or "flavor" of a report is influenced primarily by the training and orientation of the examiner. The clinician can choose from four general report-writing approaches: literary, clinical, scientific, and professional (Ownby, 1997; Tallent, 1992, 1993). Each style has unique strengths, and all have a number of liabilities. The literary approach uses everyday language, is creative, and often dramatic. Although it can effectively capture a reader's attention and provide colorful descriptions, it is often imprecise and prone to exaggeration.

The clinical approach focuses on the pathological dimensions of a person. It describes the client's abnormal features, defenses, dynamics involved in maladjustment, and typical reactions to stress. The strength of the clinical approach is that it provides information about areas in need of change and alerts a potential practitioner to likely difficulties during the course of treatment. However, such a report tends to be one-sided in that it may omit important strengths of the person. The result is likely to be more a description of a "patient" than a person. Such a maladjustment bias is a frequent difficulty in clinical psychology and results in a distorted, unrealistic view of the client. Although most clinical reports should describe a person's problem areas, these problem areas should be given appropriate emphasis in the context of the client's relevant strengths and resources.

The scientific approach to report writing emphasizes normative comparisons, tends to be more academic, and, to a lesser extent, relates to the nature of a client's pathology. The scientific style differs from the other two approaches chiefly in its reference to concepts, theories, and data. It looks at and describes test findings in an objective, factual manner. Thus, there might be frequent references to test data, normative comparisons, probability statements, and cutoff scores to be used for decision making. A scientific approach is likely to discuss the person by addressing different, often isolated, segments of personality. Thus, areas such as a client's cognitive, perceptual, and motivational abilities may be described as discrete and often unrelated functions. Although the scientific

approach is objective and factual, it has been criticized for violating the unity of personality. Many readers, particularly those from other disciplines, do not respect or empathize with scientific evaluations and perceive them as cold, distant, and overly objective. Purely data-oriented evaluations can potentially do the profession a disservice by reinforcing the view that an assessment is like a laboratory test rather than a professional consultation with a clinician. Furthermore, a focus on factual data may not address the practical decisions the client and referral source are facing.

In actual practice, it is unusual to find a pure example of a literary, clinical, or scientific report. Clinicians generally draw from all three approaches but typically emphasize one. An important part of effective report writing is the ability to evaluate the assets and limitations of each style, and to maintain a flexible orientation toward appropriately combining them. In any one report, there may be a need to use creative literary descriptions, elaborate on different pathological dimensions, or provide necessary scientific information. Again, the key is to avoid the pitfalls associated with specializing in any one of these styles and to emphasize instead their relative strengths.

Ownby (1997) stresses that the most important style to use in report writing is what he refers to as a *professional style.* This is characterized by short words that are of common usage and that have precise meanings. Grammatically, writers should use a variety of sentence constructions and lengths to maintain the reader's interest. The paragraphs should be short and should focus on a single concept. Similar concepts should be located close to one another in the report. Whereas Hollis and Donna (1979) urge writers to use short words, short sentences, and short paragraphs, the *Publication Manual of the American Psychological Association* (5th ed., 2001) recommends varying the lengths of sentences and paragraphs. The result should be a report that combines accuracy, clarity, integration, and readability.

Presenting Test Interpretations

Clinicians generally prefer to orient their reports around specific hypotheses or different relevant domains, or adhere to interpreting the data test by test. The hypothesis-oriented model focuses heavily on answering specific questions asked by the referral source. The report tends to be highly focused, well integrated, and avoids any extraneous material. For example, if a referral source asks whether person X is brain-damaged, all the interpretations based on the test data are directed toward answering whether this hypothesis is supported.

A domain-oriented report discusses the client in relation to specific topics such as cognitive abilities, interpersonal relationships, vocational abilities, or sexuality. This approach is comprehensive, indicates the client's strengths and weaknesses, and typically gives the reader a good feel for the person as a whole. The referral question is still answered but is addressed by responding to specific domains relating to the referral question. Readers tend to prefer and better comprehend integrated reports written by addressing functional domains rather than test scores. The weakness of domain-oriented reports lies in the potential to provide too much information, thus overloading the reader.

Occasionally, a report is organized by presenting the results of each test, one at a time (WAIS-III, Bender, MMPI-2, etc.). This approach clarifies the source of the data and enables the reader to understand more clearly how the clinician made his or her inferences.

It is also relatively easy for the examiner to organize the results. These advantages are offset by significant disadvantages. The emphasis on tests can distract the reader and tends to reduce the client from a person to a series of test numbers. This is reflected in that readers of reports, regardless of their theoretical or disciplinary background, do not respond well to this style of report writing (Mendoza, 2001; Tallent 1992, 1993). A test-by-test presentation also reflects a failure to integrate the data. This may be a particular issue because it is not uncommon for inconsistencies to occur among different test scores. Often, only half of all possible interpretations listed in an interpretive manual or computer narrative are actually true for a particular client. It is up to the clinician to determine which of these do or do not apply for the person. Sometimes a report writer using a test-by-test approach hedges his or her "interpretations" by using a phrase such as "Other persons with similar test profiles have the following qualities . . ." The referral source, however, doesn't want to know about *other* people but is concerned with *this* client, at this time, living in a certain context. A test-by-test interpretation, then, suggests that the practitioner has neither adequately conceptualized relevant dynamics, nor fully understood the area under investigation (Mendoza, 2001; Wolber & Carne, 1993). It also encourages the belief that an examiner is a technician who merely administers tests rather than a clinician who uses multiple sources of information to answer referral questions and help people solve problems they are facing. The existing literature is unanimous in discouraging a test-by-test style and, instead, strongly recommends an integrated, case-focused, problem-solving style (Beutler & Groth-Marnat, 2003; Groth-Marnat, in press; Kvaal, Choca, & Groth-Marnat, 2003; Mendoza, 2001; Sattler, 2002; Tallent, 1992, 1993; Wolber & Carne, 1993; Zuckerman, 2000).

Topics

There is an extremely wide range of topics or domains that clinicians may decide to discuss in their reports. These topics serve as conceptual tools that enable report writers to give form and direction to the information they are trying to communicate. The three most common topics are likely to be related to cognitive functioning, emotional functioning (affect/mood), and interpersonal relations. Many reports can be adequately organized around these three domains. Additional topics include personal strengths, vocational aptitudes, suicidal potential, defenses, areas of conflict, behavior under stress, impulsiveness, or sexuality. Often, an adequate case-focused report can be developed by describing just a few of these topics. For example, a highly focused report may elaborate on one or two significant areas of functioning, whereas a more general evaluation may discuss seven or eight relevant topics. Table 15.1 is a representative list of topics that may be considered for inclusion in an evaluation. This list is by no means complete but can provide a general guide for the wide range of possible topics from which a report writer can choose.

Deciding What to Include

The general purpose of a psychological evaluation is to provide information that will be most helpful in responding to the referral question and meeting the needs of the client. In this context, the clinician must strike a balance between providing too much

Table 15.1 Examples of general topics around which a case presentation may be conceptualized

Achievement	Emotional functioning	Psychopathology
Affect	Fixations	Rehabilitation needs
Aggressiveness	Flexibility	Rehabilitation prospects
Antisocial tendencies	Frustrations	Resistance
Anxieties	Functional impairment	Sentiments
Aptitudes	Goals	Sex
Attitudes	Hostility	Sex identity
Aversions	Identity	Sex role
Awareness	Intellectual controls	Significant others
Behavioral problems	Intellectual levels	Situational factors
Biological risk factors	Interests	Social consequences of
Cognitive functioning	Interpersonal relations	behavior
Cognitive skills	Interpersonal skills	Social role
Cognitive style	Irrational cognitions	Social structure
Competency	Lifestyle	Social support
Conflicts	Mood	Special assets
Content of consciousness	Needs	Strengths
Coping style	Outlook	Subjective feeling states
Defenses	Perception of environment	Symptoms
Deficits	Perception of self	Treatment prospects
Developmental factors	Personal consequences of	Value system
Diagnostic considerations	behavior	Vocational topics
Drives, dynamics	Placement prospects	
Emotional controls	Problem complexity	

Adapted from N. Tallent (1988). *Psychological Report Writing* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall, p. 120.

information and providing too little, and between being too cold and being too dramatic. As a rule, information should be included only if it serves to increase the understanding of the client. For example, descriptions of a client's appearance should be oriented toward areas such as his or her level of anxiety or resistance. A client might be described as hesitant in his or her approach to tasks and may say something like, "Why do I have to take all these tests anyway?" If the person was dressed in bizarre clothes and his or her hair was unkempt or dyed purple, this information might also be quite important to include. Generally, however, information regarding the types of clothing the person is wearing or color of his or her eyes or hair is not relevant.

The basic guidelines for deciding what to include in a report relate to the needs of the referral setting, background of the readers, purpose of testing, relative usefulness of the information, and whether the information describes unique characteristics of the person. After these general guidelines have been considered, the next step is to focus on and organize the information derived from the tests. For example, if a general review of aspects of personality is the purpose of the report, a clinician can look at each test to determine what information it can provide.

A further general rule is that information should focus on the client's unique method of psychological functioning. A reader is concerned not so much with how the client is

similar to the average person as in what ways he or she is different. A common error in psychological reports is the inclusion of generalized statements that are so vague, they could apply to the majority of the population. These vague, generalized statements are likely to be unconditionally accepted as applying to a person even though they are randomly selected. For example, Sundberg (1955) administered a "personality" test to a group of students and gave them all identical "interpretations" based on universal or stereotyped personality descriptions composed of 13 statements, such as:

- You have a great need for other people to like and admire you.
- You have a tendency to be critical of yourself.
- You have a great deal of unused capacity you have not turned to your advantage.
- While you have some personality weaknesses, you are generally able to compensate for them.
- At times, you have serious doubts as to whether you have made the right decision or done the right thing.

Virtually all students used in the study reported that the evaluation statements were accurate descriptions of themselves. Other studies suggest that, not only were students unable to discriminate between fictitious and genuine feedback, but they may even have preferred generalized fictitious results, particularly if they were framed within a positive context (Dies, 1972). This uncritical acceptance of test interpretations might be even further encouraged when objective-appearing, computer-generated interpretations are used (Groth-Marnat & Schumaker, 1989). W. Klopfer (1960) has referred to this uncritical acceptance of universally valid statements as the "Barnum effect," in reference to Phineas Barnum's saying, "There is a fool born every minute." Although "universal statements" may add to the "subjective" validity of the report when read by the client, such statements should be avoided in favor of stressing the person's essential uniqueness.

After the data, conclusions, and recommendations have been outlined, the next step is to decide on the manner in which to present them. This involves clear communication about the relative degree of emphasis of the results, type of report, proper use of terminology, and the extent to which the raw data will be discussed.

Emphasis

Careful consideration should be given to the appropriate emphasis of conclusions, particularly when indicating the relative intensity of a client's behavior. General summaries may be given, such as "this client's level of depression is characteristic of inpatient populations," or the relative intensity of certain aspects of a client's disorder may be more specifically discussed. To continue with the example of depression, a clinician may discuss the client's cognitive self-criticisms, degree of slowed behavior, extent of social support, level of social skills, or suicidal potential. In addition to discussing and giving the appropriate degree of emphasis to a client's pathology, his or her psychological strengths need to be compared with his or her relative weaknesses. Furthermore, the report should not discuss areas of minor relevance unless they somehow relate to the purpose of the evaluation. To achieve proper emphasis, the examiner and the referral source

must clarify and agree on the purpose of the evaluation. Only after this has been accomplished can the examiner decide whether certain information should be elaborated in-depth, briefly mentioned, or deleted.

When clinicians present their conclusions, it is essential that they indicate their relative degree of certainty. Is a specific conclusion based on an objective fact, or is the clinician merely presenting a speculation? For example, the statement "John scored in the dull normal range of intelligence" is an objective fact. However, even in this case, examiners may want to give the standard error of measurement to provide an estimate of the probable range of scores. If only mild supporting data is available or if clinicians are presenting a speculation, phrases such as "it appears . . .", "tends to . . .", or "probably . . ." should be used. This is especially important when clinicians are attempting to predict a person's behavior, because the predicted behavior has not yet been observed. It may be useful for clinicians to indicate that their predictions cannot be found directly in the tests themselves, but rather represent inferences that have been made based on the test data. There should be a clear distinction between what the client did, and what he or she anticipates doing. If a statement made in a report is a speculation, it should be clearly indicated that the statement has only a moderate or small degree of certainty. Whenever a speculation is included, it should be relevant to the referral question.

Improper emphasis can reflect an incorrect interpretation by the examiner, and this misinterpretation is then passed down to the reader. Clinicians sometimes arrive at incorrect conclusions because their personal bias results in selective perception of the data. Thus, clinicians can develop an overly narrow focus with which they overlook potentially relevant data. Personal bias may result from factors such as a restrictive theoretical orientation, incorrect subjective feelings regarding the client, or an overemphasis on pathology. Inaccurate conclusions can also result from attempts to please the referral source or from interpretations based on insufficient data. The reader may also be likely to misinterpret the conclusions if the report is generally overspeculative or if speculations are not specified as such but, rather, are disguised as assertions. If speculations are overly assertive, this may not only lead the reader to develop incorrect conclusions, but also the report may become overly authoritative and dogmatic, perhaps leading readers to become irritated and skeptical.

Misinterpretations can also result from vague and ambiguously worded sentences that place incorrect or misleading emphasis on a client's behavior. A statement such as "the client lacks social skills" is technically incorrect because the client must have some social skills, although these skills may be inadequate. A more correct description would be to state that the client's social skills are "poorly developed" or "below average." Likewise, a statement such as "the client uses socially inappropriate behavior" is subject to myriad interpretations. This could be rephrased to include more behaviorally oriented descriptions, such as "frequently interrupts" or "would often pursue irrelevant tangents."

One technique of emphasizing results is to place the most relevant sections in boldface or italics. For example, the major identified symptoms, most important findings, and the major recommendations could all be placed in boldface. This enables persons reading the report to more easily absorb the most salient features. In addition, they can easily relocate major points that have been made. However, this technique should be used sparingly because readers can become easily saturated with too much boldface

print. This means that, instead of placing entire paragraphs or sentences in boldface, only key phrases should be given this method of emphasis.

The areas, extent, and method of emphasis significantly contribute to the conclusions of a report. However, responsibility for a report's conclusions rests on the clinician. This responsibility should not and cannot be transferred to the tests themselves. To take this a step further, decisions made about a person should never be in the hands of tests, which may even have questionable validity in certain contexts. Rather, conclusions and decisions regarding people should always be in the hands of responsible persons. Thus, the style of emphasizing results should reflect this. Phrases such as "test results indicate . . ." may give the impression that the examiner is trying to hide behind and transfer responsibility for his or her statements onto the tests. Not only is this not where the responsibility should be, but the reader may develop a lack of confidence in the clinician. If clinicians feel uncertain about a particular area, they should either be clear about this uncertainty or, if they cannot personally stand by the results, exclude the results from the report.

Use of Raw Data

When writing the impressions and interpretation section, a report writer should generally avoid adhering too closely to the raw data. For certain purposes, however, it may be useful to include raw data or even to describe the tests themselves. Test descriptions allow untrained persons to know specific behaviors the client engaged in rather than merely the final inferences. As a result, consumers of reports rate the inclusion of behavioral referents quite favorably (Finn et al., 2001). For example, a report may include a description such as "Mr. A had an average level of recall for short-term visual information, as indicated by his being able to accurately recall and reproduce five out of a possible nine geometric designs that he had previously worked with for five minutes." This sentence provides a more behaviorally referenced description than one like "Mr. A had an average level of recall as measured on the Bender memory." Thus, a test description is apt to give the reader a more in-depth, precise, and familiar reference regarding the subject's abilities. In addition to the test descriptions themselves, test responses can serve to make the description behavior specific and to balance high-level abstractions with concrete responses. For example, a clinician might discuss a client's impulsiveness and include illustrative items on the MMPI-2 (items adapted from actual MMPI-2 items), such as:

> During one period when I was a child, I used to shoplift. (True)
>
> At times, I have found it nearly impossible to keep from stealing. (True)

In discussing the same issue, a clinician could also include a portion of a TAT story that illustrates a similar point:

> . . . so he took the violin and, without even thinking about it, threw it into the fire and ran.

It is crucial to stress that the purpose of providing raw data and behavioral descriptions is to enrich and illustrate the topic and not to enable the reader to follow

the clinician's line of reasoning or document the inferences that have been made. In developing inferences, clinicians must draw on a wide variety of data. They cannot possibly discuss all the patterns, configurations, and relationships they used to come to their conclusions. Any attempt to do so would necessarily be overly detailed, cumbersome, and incomplete. Statements such as, "In considering the pattern of elevated Scales 4 and 9 on the MMPI, it is safe to conclude . . ." are unnecessary and rarely contribute to a report's overall usefulness. In certain types of reports, such as those for legal purposes, it might be helpful to include some raw data, not so much to repeat the thinking process of the clinician, but more to substantiate that the inferences are data based, to provide a point of reference for discussing the results, and to indicate what assessment procedures were used.

Terminology

Several arguments have been made in determining whether to use technical or nontechnical language in psychological reports. It might be argued that technical terminology is precise and economical, increases the credibility of the writer, and can communicate concepts that are impossible to convey through nontechnical language. However, a number of potential difficulties are often encountered with the use of technical language. One of the more frequent problems involves the varying backgrounds and levels of sophistication of the persons reading the report. The most frequent readers of reports include teachers, administrators, judges, attorneys, psychiatrists, nonpsychiatric physicians, and social workers. Increasingly, the clients themselves have access to and read the reports (Harvey, 1997). Thus, many, if not most, consumers of reports do not have the necessary background to interpret technical terminology accurately. Even psychologists with different theoretical persuasions may be apt to misinterpret some of the terms. Take, for example, the differing uses of *ego* by Freud, Jung, and Erikson. Also, the term *anxiety* might have several different categories of use. Although technical words can undoubtedly be precise, their precision is helpful only in a particular context and with a reader who has the proper background. Generally, reports are rated as more effective when the material is described in clear, basic language (Finn et al., 2001; Harvey, 1997; Ownby, 1990, 1997; Sandy, 1986; Tallent, 1993). Even among readers who have the proper background to understand technical terms, many prefer a more straightforward presentation (Ownby, 1990, 1997; Tallent, 1992, 1993). Technical terms also run the danger of becoming nominalisms in which, by merely naming the phenomenon, persons develop an illusory sense of understanding more than is actually the case. Terms such as *immature* or *sadistic* cover a great deal of information because they are so general, but they say nothing about what the person is like when he or she is behaving in these maladaptive ways. They also do not adequately differentiate one person from the next and are frequently ambiguous. Furthermore, technical terms are often used inappropriately (e.g., a person who is sensitive and cautious in interpersonal relationships is labeled *paranoid,* or *compulsive* is used to describe someone who is merely careful, conscientious, and effective in dealing with details).

W. Klopfer (1960) provides an excellent and still-relevant rationale for using basic English rather than technical terminology. First, and perhaps most important, the use

of basic English allows the examiner, through his or her report, to communicate with and affect a wide audience. This is particularly important because the number and variety of persons who read reports is much greater now than 20 or 30 years ago. Furthermore, basic English is more specific and descriptive of an individual's uniqueness, whereas technical terms tend to deal with generalities. Terms such as *sadomasochistic* and *hostile* do not provide essential information about whether the person is assaultive or suicidal. Finally, the use of basic English generally indicates that the examiner has more in-depth comprehension of the information he or she is dealing with and can communicate this comprehension in a precise, concrete manner. Klopfer stresses that any description found in a psychological report should be comprehensible to any literate person of at least average intelligence. In contrast, psychologists have been found to write in a more technical and complex level when compared with the average client (Harvey, 1997). The first four are examples of translating technical concepts into basic English (Klopfer, 1960):

> "Hostility toward the father figure" becomes "the patient is so fearful and suspicious of people in positions of authority that he automatically assumes an aggressive attitude toward them, being sure that swift retaliations will follow. He doesn't give such people an opportunity to demonstrate their real characteristics since he assumes they are all alike."
>
> "The patient projects extensively" becomes "the patient has a tendency to attribute to other people feelings and ideas originating within himself regardless of how these other people might feel."
>
> "The defenses the patient uses are . . ." becomes "the methods characteristically employed by the patient for reducing anxiety are . . ."
>
> "Empathy" becomes "the patient can understand and sympathize with the feelings of others, since she finds it relatively easy to put herself in their place." (pp. 58–60)
>
> "The client is hostile and resistant" may be changed to include a behavioral description; "when the client entered the room she stated, 'My Dad said I had to come and that's the only reason I'm here' " or "later on in the testing she made several comments such as 'This is a stupid question.' "

The general principle involved in the preceding examples is to translate high-level abstract terms into basic English that provides useful, concrete behavioral descriptions.

Ownby (1990, 1997) recommends combining any conclusion or generalization with specific behaviors or test observations. Recommendations should also be directly linked with the relevant behaviors/generalizations, either in the same place or in the recommendations section. For example, instead of saying a client is "depressed," a writer might state, "The client's behavior, which included self-criticism and occasional crying, suggested he was depressed." Linking generalizations with clear concrete descriptions tends to create reports that are perceived to be relatively credible and persuasive (Ownby, 1990, 1997). If this process is followed, descriptions will be less subject to misinterpretation, less ambiguous, and more likely to convey the unique characteristics of the client. Although abstract technical terms can at times be important components of a psychological report, they should be used sparingly and only when clearly appropriate. This particularly means carefully considering the background of the persons who will read the report. Sandy (1986) even recommends having the clinician collaborate with the relevant recipients of the report, including the client, so that the final report is descriptive rather

than interpretive and the readers are not passive recipients of the "higher" wisdom of the psychologist.

Content Overload

There are no specific rules to follow in determining how much information to include in a report. A general guideline is to estimate how much information a reader can realistically be expected to assimilate. If too many details are given, the information may begin to become poorly defined and vague and, therefore, lack impact or usefulness. When clinicians are confronted with a great variety of data from which to choose, they should not attempt to include it all. A statement such as "The client's relative strengths are in abstract reasoning, general fund of knowledge, short-term memory, attention span, and mathematical computation" is likely to overload the reader with too many details. The clinician should instead adequately develop each of the various points and focus on the areas that are most relevant to the purpose of the report.

Feedback

During the earlier days of psychological assessment, examiners often kept the results of psychological assessments carefully concealed from the client. There was often an underlying belief that the results were too complex and mysterious for the client to adequately understand. In contrast, current practices are to provide the client with clear, direct, and accurate feedback regarding the results of an evaluation (S. Ackerman, Hilsenroth, Baity, & Blagys, 2000; Finn & Tonsager, 1997; Lewak & Hogan, 2003; K. Pope, 1992).

The change toward providing feedback to clients has been motivated by several factors. First, regulations have supported a growing list of consumer rights, including the right to various types of information. Second, it might be perceived as a violation if the client did not receive feedback regarding the results of testing after he or she had been subjected to several hours of assessment. Even the most secure of clients might easily feel uncomfortable knowing a report with highly personal information might be circulated and used by persons in power to make decisions about the client's future. Such practices could understandably result in suspicion and irritation on the part of the public. Third, examiners cannot safely assume that the original referral source will provide feedback to the client. Even if the referral source does provide feedback, there is no guarantee that the information will be provided in an appropriate manner. Thus, the responsibility for providing feedback is ultimately on the clinician. Finally, there is increasing evidence that providing clients with test feedback can result in significant therapeutic benefits (S. Ackerman et al., 2000; Finn & Tonsager, 1992; Gass & Brown, 1992).

The extent to which a clinician providing feedback will allow the client to actually read all or portions of the report varies. The rationale for allowing the client to actually read the report is that doing so enables the client to experience the product of assessment in a direct manner. It also enables a practitioner to explain any areas that are unclear. A significant difficulty is that the client might misinterpret various portions of the report, especially IQ scores and diagnosis. For this reason, most clinicians paraphrase and elaborate on selected portions of the report. This increases the likelihood that clients will

readily understand the most important material and will not be overloaded with too much content.

The likelihood of providing effective feedback can be enhanced by following several guidelines. Initially, the rationale for assessments should be explained and any misconceptions should be corrected. One particularly important misconception is that sometimes clients mistakenly fear that the purpose of assessment is to evaluate their sanity. Practitioners must also select the most essential information to be conveyed to the client. To a large extent, this involves clinical judgment. Important considerations include the client's ego strength, life situation, stability, and receptiveness to different types of material. Typically, three to four general and well-developed areas represent an optimum amount of information. The information that is provided should be carefully integrated into the overall context of the person's life. This integration might be enhanced by providing concrete behavioral examples, reflecting on aspects of the client's behavior, referring to relevant aspects of the client's history, or paraphrasing and expanding on a client's self-descriptions. A useful technique is to have the client evaluate the relevance and accuracy of the information. The client might also be asked to give his or her own examples of the trait or pattern of behavior described in the report. Such collaboration with the client helps the clinician to determine how well the client has understood the feedback. Underlying any feedback should be an attempt to provide the information in a clear, intelligible manner. Commonplace language should be used instead of psychological jargon. It is also important to take into account the client's level of intelligence, education, vocabulary, and level of psychological sophistication. Feedback should be not only a neutral conveyance of data but also a clinical intervention. The information should provide the client with new perspectives and options and should aid in the client's own problem solving.

One possibility is to prepare a personalized report designed specifically for the client. This forces the practitioner to write in a clear, straightforward style. Such reports are more likely to emphasize adaptation rather than pathology. In addition, clear recommendations tend to be emphasized. The optimal communication style is an informal letter written to and for the client ("I am writing to communicate the findings of our psychological assessment . . ."). There are currently available a number of computerized reports directed toward providing the client with feedback. There also seems to be a trend for additional resources to include interpretations directed toward the client such as Lewak, Marks, and Nelson's (1990) *Therapist Guide to the MMPI & MMPI-2: Providing Feedback and Treatments* and Finn's (1996) *Manual for Using the MMPI-2 as a Therapeutic Intervention.* These sources should not be seen as substitutes for a dynamic interaction with a client, but as adjuncts for enhancing this process.

FORMAT FOR A PSYCHOLOGICAL REPORT

Although no single, agreed-on format exists, every report should integrate old information as well as provide a new and unique perspective on the person. Old information should include identifying information (name, birth date, etc.), reason for referral, and relevant history. New information should include assessment results, impressions, summary/conclusions, and recommendations. At the top of the report, practitioners should

indicate its confidential nature by writing "Confidential Psychological Evaluation." A suggested outline follows:

Name:
Age (date of birth):
Sex:
Ethnicity:
Date of Report:
Name of Examiner:
Referred by:
I. Referral Question
II. Evaluation Procedures
III. Behavioral Observations
IV. Background Information (relevant history)
V. Test Results
VI. Impressions and Interpretations
VII. Summary and Recommendations

Although this outline represents a frequently encountered format, there are many variations. Some practitioners prefer to include the client's marital status, occupational status, and handedness (for neuropsychological reports) at the top of the report along with the other demographic information. Other practitioners prefer to exclude the test results section or include additional sections on diagnosis, case formulation, or summary. Still others like to include subheadings in the Impressions and Interpretation section (cognitive functioning and ideation, coping style, affect/mood/emotional control). Sometimes it might be more appropriate to eliminate all or most of the headings and write the report directly to the referral source in a letter format ("Dear Dr. Jones: . . ."). The sample reports included later in this chapter have purposely been chosen to demonstrate a variety of different formats in diverse styles and contexts. Each practitioner needs to develop both the format and style that most effectively meet his or her client's and referral source's needs. In addition, different assessment contexts require different styles and areas of focus.

Referral Question

The Referral Question section provides a brief description of the client and a statement of the general reason for conducting the evaluation. In particular, this should include a brief description of the nature of the problem. If this section is adequately completed, it should give an initial focus to the report by orienting the reader to what follows and to the types of issues that are addressed. This section should begin with a brief, orienting sentence that includes essential information about the client ("Mr. Smith is a 35-year-old, White, married male with a high school education who presents with complaints of depression and anxiety"). Such a sentence clearly and succinctly introduces the client. A prerequisite for this section is that the clinician has developed an

adequate clarification of the referral question. The purpose of testing should be stated in a precise and problem-oriented manner. Thus, phrases such as "the client was referred for a psychological evaluation" or "as a requirement for a class project" are inadequate because they lack focus and precision. It is helpful to include both the specific purpose of the evaluation and the decisions facing the referral source.

Examples of possible reasons for referral include:

- Intellectual evaluation: routine, intellectually disabled (retarded), gifted.
- Differential diagnosis, such as the relative presence of psychological difficulties (i.e., memory problems caused by depression) versus organic impairment (i.e., memory problems because of the early stages of Alzheimer's disease).
- Assessment of the nature and extent of brain damage.
- Evaluation as a component of, and to provide recommendations for, vocational counseling.
- Evaluation of appropriateness for, possible difficulties encountered in, and optimal approach to, psychotherapy.
- Personal insight regarding difficulties with interpersonal relationships.
- Evaluation as an aid in client placement.

These represent general referral questions that, in actual situations, would require further clarification, especially regarding the decisions facing the referral source (see Armengol, 2001). The key should be to find out what the referring person really wants from the report. This may require reading beneath the surface of the referral question(s) and articulating possible hidden agendas and placing the referral question into a wider context than the presenting problem. In some cases, it may be necessary to educate the referral source regarding the strengths as well as the limitations of psychological testing. This may even lead to recommending that the person not be tested. An effective referral question should accurately describe the client's and the referral source's current problems.

After the referral questions have been clarified and outlined, they can be addressed throughout the rest of the report. It is usually helpful to succinctly reiterate and summarize the answers to the referral questions toward the end of the report. Some clinicians prefer to carefully note each of the referral questions that have been made at the beginning of the report, and then bullet/number succinct answers to each of these questions in the Summary and Recommendations. Such a procedure is user friendly, provides succinct answers to the questions, and allows for symmetry and closure to the report.

Evaluation Procedures

The report section that deals with evaluation procedures simply lists the tests and other evaluation procedures used but does not include the actual test results. Usually, full test names are included along with their abbreviations. Later in the report, the abbreviations can be used, but the initial inclusion of the entire name provides a reference for readers who may not be familiar with test abbreviations. For legal evaluations or other occasions in which precise details of administration are essential, it is important to include

the date on which different tests were administered and the length of time required to complete each one. For most routine evaluations, however, this degree of detail is not recommended. It may also be important to include whether a clinical interview or mental status examination was given and, if so, the degree of interview structure and the amount of time required for the interview or examination. Evaluation procedures may not necessarily be restricted to testing and interviews with the client. Often, evaluation includes a review of relevant records such as medical reports, nursing notes, military records, police records, previous psychological or psychiatric reports, or educational records. Additional material might come from interviews with individuals such as spouses, children, parents, friends, employers, physicians, lawyers, social workers, or teachers. If any of these sources are used, their dates and, if relevant, who wrote them should be included. This section might end with a statement summarizing the total time required for the evaluation.

Behavioral Observations

A description of the client's behaviors can provide insight into his or her problem and may be a significant source of data to confirm, modify, or question the test-related interpretations. These observations can be related to a client's appearance, general behavioral observations, or examiner-client interaction. Descriptions should be tied to specific behaviors and should not represent a clinician's inferences. For example, instead of making the inference that the client was "depressed," it is preferable to state that "her speech was slow and she frequently made self-critical statements such as 'I knew I couldn't get that one right.' "

Relevant behavioral observations made during the interview include physical appearance, behavior toward the task and examiner, and degree of cooperativeness. A description of the client's physical appearance should focus on any unusual features relating to facial expressions, clothes, body type, mannerisms, and movements. It is especially important to note any contradictions, such as a 14-year-old boy who acts more like an 18-year-old or a person who appears dirty and disheveled but has an excellent vocabulary and high level of verbal fluency. The behaviors the client expresses toward the test material and the examiner often provide a significant source of information. These may include behaviors that reflect the person's level of affect, manifest anxiety, presence of depression, or degree of hostility. The client's role may be as an active participant or generally passive and submissive; he or she may be very much concerned with his or her performance or relatively indifferent. The client's method of problem solving is often a crucial area to note, and it may range from careful and methodical to impulsive and disorganized. It is also important to pay attention to any unusual verbalizations that the client makes about the test material. The level of cooperation expressed by the client should be a factor in assessing the validity of the test results. This is especially important for intelligence and ability tests, because a prerequisite is that the client be alert and attentive, and put forth his or her best effort. It may also be important to note events before testing, such as situational crises, previous night's sleep, or use of medication. If there are situational factors that may modify or bring into question the test's validity, they should be noted with statements such as, "The test results should be viewed with caution because . . ." or "The degree of maladjustment indicated on the test scores may

represent an exaggeration of the client's usual level of functioning due to conditions surrounding the test administration." Often, the most important way to determine test validity in relationship to the client is through a careful look at the client's behaviors relating to the tests and his or her life situation before testing.

Sattler (1992) has developed a Behavior and Attitude Checklist comprising 10 major categories that can be rated on a seven-point scale (see Table 15.2). The examiner may wish to use this checklist as a tool to help focus on areas that might be significant to mention or discuss. It is important to emphasize that other crucial behaviors that are not covered in the checklist may occur and still require discussion.

Behavioral observations should usually be kept concise, specific, and relevant. If a description does not allow for some insight about the person or demonstrate his or her uniqueness, it should not be included. Thus, if a behavior is normal or average, it is usually not important to discuss other than to briefly mention that the person had, for example, an average level of cooperation, alertness, or anxiety. The focus, then, should be on those client behaviors that create a unique impression. The relative length of this section varies from a few brief sentences to considerably longer depending on the amount of relevant information the clinician has noticed. The relative importance of this section in relationship to the overall report is, likewise, extremely varied. Sometimes, this section can be almost as important as the test results, whereas at other times, it might consist of a few minor observations.

Clinicians who prefer behavioral assessment procedures might wish to emphasize the behavioral observation section by providing more in-depth descriptions of relevant antecedents. In addition, consequent events surrounding the problem behavior itself might be evaluated in relationship to their onset, duration, frequency, and intensity. Specific strategies of behavioral assessment include narrative descriptions, interval recording, event recording, ratings recordings, and self-report inventories (see Chapter 4).

Some examiners may wish to summarize information from a Mental Status Examination in the Behavioral Observations section. In these cases, there is necessarily a movement away from concrete descriptions of behaviors to inferences about these behaviors. For example, a clinician may infer, based on behavioral observations, that the client was oriented to time and place. Additional categories might include verbalizations, psychomotor activity, affect, thought processes/contents, and insight/judgment (see section on the Mental Status Examination in Chapter 3).

Another exception to adhering exclusively to concrete behavioral descriptions is that, at the end of the Behavioral Observations section, it is customary and appropriate to include a statement indicating the validity of the assessment procedures. For example, it might state something like: "Given the consistency and detail of the client's responses, the client's high level of motivation, and validity indicators on the MMPI-2, the assessment appears to be an accurate assessment of this person's current level of functioning."

Background Information (also Referred to as Relevant History)

The write-up of a client's background information should include aspects of the person's history that are relevant to the problem the person is confronting and to the interpretation of the test results. The history, along with the referral question, should also

Table 15.2 Behavior and Attitude Checklist

Client's name: _____ Examiner: _____

Age: _____ Date of report: _____

Test(s) administered: _____ Date of examination: _____

IQ: _____ Grade: _____

Instructions: Place an X on the appropriate line for each scale.

I.	*Attitude toward examiner and test situation:*		
	1. cooperative	____ : ____ : ____ : ____ : ____ : ____ : ____	uncooperative
	2. passive	____ : ____ : ____ : ____ : ____ : ____ : ____	aggressive
	3. tense	____ : ____ : ____ : ____ : ____ : ____ : ____	relaxed
	4. gives up easily	____ : ____ : ____ : ____ : ____ : ____ : ____	does not give up easily
II.	*Attitude toward self:*		
	5. confident	____ : ____ : ____ : ____ : ____ : ____ : ____	non confident
	6. critical of own work	____ : ____ : ____ : ____ : ____ : ____ : ____	accepting of own work
III.	*Work habits:*		
	7. fast	____ : ____ : ____ : ____ : ____ : ____ : ____	slow
	8. deliberate	____ : ____ : ____ : ____ : ____ : ____ : ____	impulsive
	9. thinks aloud	____ : ____ : ____ : ____ : ____ : ____ : ____	thinks silently
	10. careless	____ : ____ : ____ : ____ : ____ : ____ : ____	neat
IV.	*Behavior:*		
	11. calm	____ : ____ : ____ : ____ : ____ : ____ : ____	hyperactive
V.	*Reaction to failure:*		
	12. aware of failure	____ : ____ : ____ : ____ : ____ : ____ : ____	unaware of failure
	13. works harder after failure	____ : ____ : ____ : ____ : ____ : ____ : ____	gives up easily after failure
	14. calm after failure	____ : ____ : ____ : ____ : ____ : ____ : ____	
	15. apologetic after failure	____ : ____ : ____ : ____ : ____ : ____ : ____	not apologetic after failure
VI.	*Reaction to praise:*		
	16. accepts praise gracefully	____ : ____ : ____ : ____ : ____ : ____ : ____	accepts praise awkwardly
	17. works harder after praise	____ : ____ : ____ : ____ : ____ : ____ : ____	retreats after praise
VII.	*Speech and language:*		
	18. speech poor	____ : ____ : ____ : ____ : ____ : ____ : ____	speech good
	19. articulate language	____ : ____ : ____ : ____ : ____ : ____ : ____	inarticulate language
	20. responses direct	____ : ____ : ____ : ____ : ____ : ____ : ____	responses vague
	21. converses spontaneously	____ : ____ : ____ : ____ : ____ : ____ : ____	only speaks when spoken to
	22. bizarre language	____ : ____ : ____ : ____ : ____ : ____ : ____	reality-oriented language
VIII.	*Visual-motor:*		
	23. reaction time slow	____ : ____ : ____ : ____ : ____ : ____ : ____	reaction time fast
	24. trial-and-error	____ : ____ : ____ : ____ : ____ : ____ : ____	careful and systematic
	25. skillful movements	____ : ____ : ____ : ____ : ____ : ____ : ____	awkward movements
IX.	*Motor:*		
	26. defective motor coordination	____ : ____ : ____ : ____ : ____ : ____ : ____	good motor coordination
X.	*Overall test results:*		
	27. reliable	____ : ____ : ____ : ____ : ____ : ____ : ____	unreliable
	28. valid	____ : ____ : ____ : ____ : ____ : ____ : ____	invalid

Source: From *Assessment of Children* (3rd ed.), p. 92, by J. M. Sattler, 1992, San Diego: Author. Copyright 1992 by Jerome M. Sattler, Publisher. Reprinted by permission.

place the problem and the test results into the proper context. In accomplishing these goals, the clinician does not need to include a long, involved chronology with a large number of details, but rather should be as succinct as possible. Some practitioners even urge that the background information section be kept to one concise paragraph, particularly in medical settings where there is considerable emphasis on conciseness. In selecting which areas to include and which to exclude, a clinician must continually evaluate these areas in relationship to the overall purpose of the report. It is difficult to specify precise rules because each individual is different. Furthermore, each clinician's own personal and theoretical orientation alters the types of information he or she feels are significant. Whereas one clinician may primarily describe interpersonal relationships, another may focus on intrapsychic variables, birth order, early childhood events, or details about the client's present situation and environment. The key is to maintain a flexible orientation so that the interviewer is aware of the most significant elements in the client's life. In general, the end product should include a good history of the problem, along with areas such as important life events, family dynamics, work history, personal interests, daily activities, and past and present interpersonal relationships (see Table 3.1 in Chapter 3).

When describing a client's background, it is important to specify where the information came from ("The client reported that . . ."). This is particularly essential when there may be some question regarding the truth of the client's self-reports or when the history has been obtained from multiple sources.

Usually, a history begins with a brief summary of the client's general background. This can be followed by sections describing family background, personal history, medical history, history of the problem, and current life situation.

The extent to which a clinician decides to pursue and discuss a client's family background is subject to a great degree of variability. The primary purpose of such information is to help determine causal factors, what variables might help maintain relevant behaviors, and the extent to which the family should be used as either a focus of systemic intervention or as social support. At a minimum, a brief description of the client's parents is warranted; this may include whether they are separated/divorced and alive/deceased, and their socioeconomic level, occupation, cultural background, and health status. Sometimes, it is important to include information about the emotional and medical backgrounds of parents and close relatives, because certain disorders occur with greater frequency in some families than in the overall population. A description of the general atmosphere of the family is often helpful, including the client's characteristic feelings toward family members and his or her perceptions of their relationships with each other. Descriptions of common family activities and whether the family lived in an urban or a rural environment might also be included. If one or both parents died while the client was young, the clinician can still discuss the speculations the client has about his or her parent(s) and can describe the significant persons for the client as he or she was growing up.

The client's personal history can include information from infancy, early childhood, adolescence, and adulthood. Each stage has typical areas to investigate and problems to be aware. The information from infancy usually either represents vague recollections or is secondhand information derived from parents or relatives. Thus, it may be subject to a great deal of exaggeration and fabrication. If possible, it may be helpful to have

details verified by additional sources, such as through direct questioning of parents or examination of medical records. The degree of contact with parents, family atmosphere, and developmental milestones may all be important areas to discuss. Because physical and psychological difficulties often are related and occur simultaneously, including a client's early medical history is sometimes helpful. The most significant tasks during childhood are the development of peer relationships and adjustment to school. What was the quality of the client's early friendships? How much time did the client spend with others? Were there any fights or rebellious acting out? Was the client basically a loner, or did he or she have a large number of friends? Did the person join clubs and have group activities, hobbies, or extracurricular interests? In the academic area, it may be of interest to note the usual grades, best or worst subjects, and whether the client skipped or repeated grades. Furthermore, what was his or her relationship with parents, and did the parents restrict activities or allow relative freedom? During adolescent years, clients typically face further academic, psychological, and social adjustments to high school. Of particular importance are their reactions to puberty and early heterosexual relationships. Did they have difficulties with sex role identity, abuse drugs or alcohol, or rebel against authority figures? The adult years center around occupational adjustment and establishing marital and family relationships. During early adulthood, what were clients' feelings and aspirations regarding marriage? What were their career goals? Did they effectively establish independence from parents? As adulthood progressed, were there any significant changes in the quality of their close relationships, employment, or expression of sexuality? What activities did they engage in during their leisure time? As clients age, they face challenges in adapting to their declining abilities and limitations, and developing a meaningful view of their lives.

Although the personal history can help place the problem in its proper context and explain certain causative factors, it is usually essential to spend some time focusing directly on the problem itself. Of particular importance are the initial onset and the nature of the symptoms. From the time the client first noticed these symptoms, have there been any changes in frequency, intensity, or expression? If a formal diagnosis will be made, it is particularly important to have a clear description of symptom patterns to substantiate such a diagnosis. It might also be important to determine whether there were any previous attempts at treatment, and if so, what was the outcome? In some reports, the history of the problem is the longest and most important part of the history section.

The family and personal histories usually reveal information relating to the predisposing cause of a client's difficulties, whereas the history of the problem often provides an elaboration of the precipitating and reinforcing causes. To complete this picture, the clinician also has to develop a sense of the factors currently reinforcing the problem. This requires information relating to the client's life situation. Significant areas may be the client's life stresses, including changes that he or she is confronting. In addition, what are the nature of and resources provided by his or her family and work relationships? Finally, it is important to understand the alternatives and decisions that the client is facing.

Sometimes, an evaluation needs to assess the possible presence and nature of organic impairment. In many cases, the history is of even greater significance than test results; and, often, the most valuable information a psychologist can provide to a referring medical practitioner is a thorough history. Thus, the history needs to be complete

and must address a number of areas that are not ordinarily covered in personality evaluations. Several interview aids have been commercially developed to help ensure that most relevant areas are covered (see Chapter 12). If the person reports having had a head injury, it is important to note the length of time the client was unconscious (if at all), whether he or she actually remembers getting hit, the last memory before the injury, and the first thing he or she clearly remembers following the injury. In all neuropsychological assessments, a crucial area is to establish the person's premorbid level of functioning. This may mean obtaining information on his or her grade point average in high school or college, sending for any relevant records (e.g., previous IQ results), and determining previous highest level of employment and personal interests or hobbies. Often, it may be necessary to verify the client's previous level of functioning from outside sources, such as from parents or employers. In determining the probable cause of brain impairment, it may be difficult to rule out other possibilities, such as exposure to toxic substances, strokes, high fevers, or other episodes of head trauma. Areas of current functioning that need to be addressed might include memory problems, word-finding difficulties, weakness on one side of the body, alterations in gait, loss of consciousness, and unusual sensations. Previous assessments with CT/MRI scans, EEGs, or neurological physical exams would also be important to obtain. Even though these medical records might be able to identify the site and size of a lesion, it is still the work of the psychologist to describe what the person is doing as a result of these lesions. It might also be important to obtain current or past information regarding drug intake, especially recent alterations in prescriptions, because these might affect psychological functioning. The interview data and neuropsychological test results from a psychologist should ideally be combined with and complement medical records, such as CT scans and neurological exams. Although the preceding topics are by no means exhaustive, they represent some of the more important areas to consider when taking a history related to possible neuropsychological deficit.

Although the quantity of such information may seem immense, the history format described here is only a general guideline. At times, it may be appropriate to ignore many of the areas mentioned earlier and focus on others. In condensing the client's history into the report, it is important to avoid superfluous material and continually question whether the information obtained is relevant to the general purpose of the report. Again, some practitioners might prefer to restrict the length of this section to a single paragraph. Furthermore, it is typically not useful to include material that is already in the possession of the referral source. This may be perceived as needless duplication and could result in the history section's becoming needlessly long.

Test Results

For certain reports, it may not be necessary to list test scores. Some practitioners even prefer to completely exclude actually giving test scores because it might give the impression that the report is too data/test oriented. However, it is usually recommended that, at some point, test scores be included, especially in legal reports or when professionals who are knowledgeable about testing will read the report. If practitioners feel that including this section in the report itself is too test/data oriented or serves to unnecessarily "clutter" the report, the test data, perhaps including the actual profiles,

might be inserted in an appendix. This could then be noted and the reader referred to the appendix in the evaluation procedures section.

If actual test scores are included, standard (rather than raw) scores should be the mode of presentation. Referral sources have consistently indicated that percentiles are preferred over other types of standard scores (Finn et al., 2001). Because various tests use somewhat different types of standard scores, it is recommended that each set of test scores include both the standard score and percentiles. Clinicians may also wish to indicate the relative magnitude of the relevant scores ("Very high," "High," etc.) or whether the scores exceed some clinically meaningful cutoff.

Intelligence test scores are traditionally listed first and, for the Wechsler scales, should include IQ scores (Verbal, Performance, Full Scale IQ), index scores, and subtest scaled scores. Subtests that have been found to be significant strengths should be indicated with an "S" next to the subtest score and significant weaknesses should be indicated with a "W." This is often followed by other cognitive test results such as the Bender or Wechsler Memory Scale-III. Bender results can simply be summarized by a statement such as: "Empirically not in the organic range, although there were difficulties organizing the designs and frequent erasures." MMPI-2/MMPI-A results are often listed in the order in which they appear on the profile sheet. Objective personality tests (MMPI-2, MMPI-A, CPI) should always be referred to by their standardized (usually *T*) scores and not their raw scores. Whereas it is fairly straightforward to list the objective and intelligence test scores, it is considerably more difficult to adequately describe the scores on projective tests. The Rorschach summary sheet can be included, but the results from projective drawings and the TAT are usually omitted. Should a clinician wish to summarize projective drawings, a brief statement is usually sufficient, such as "Human figure drawings were miniaturized and immature, with the inclusion of two transparencies." Likewise, TAT "scores" can be summarized by a brief statement of the strongest needs and press, and a mention of the most common themes encountered in the stories.

Impressions and Interpretations (also Referred to as Discussion)

This section can be considered the main body of the report. It requires that the main findings of the evaluation be presented in the form of integrated hypotheses. The areas discussed and the style of presentation vary according to the personal orientation of the clinician, the purpose of testing, the individual being tested, and the types of tests administered. As emphasized previously, assessment data should be organized according to different integrated topics or presented as a chronological narrative of the person. In contrast, a test-by-test presentation is strongly discouraged. To organize the information from an assessment, W. Klopfer (1960) recommends using a grid with the topics for consideration in the left column (derived from Table 15.1) with the assessment results in the top row. This enables the practitioner to extract essential findings from the data and list them in the appropriate box where the topic and the method of assessment intersect. When actually writing the Impressions and Interpretations section of the report, the clinician can then review all findings in a particular topic and summarize them on the report. An example of such a grid is given in Table 15.3. The list of assessment

Table 15.3 Sample grid of assessment domains by tests administered

	Evaluation Procedures				
Topics	Interview	WAIS-III	MMPI-2	BDI	Rorschach
Validity of results					
Cognitive functioning					
Emotional controls					
Interpersonal relations					
Diagnostic impression					
Recommendations					

Source: Adapted from Klopfer (1960), *The Psychological Report,* New York: Grune & Stratton.

methods is dependent on which tests the examiner administered, but the topics can be chosen and arranged according to areas the clinician would like to focus on.

All inferences made in the Impressions and Interpretations section should be based on an integration of the test data, behavioral observations, relevant history, and additional available data. The conclusions and discussion may relate to areas such as the client's overt behavior, self-concept, family background, intellectual abilities, emotional difficulties, medical disorders, school problems, or interpersonal conflicts. A client's intellectual abilities often provide a general frame of reference for a variety of personality variables. For this reason, a discussion of the client's intellectual abilities usually occurs first. Although this should include a general estimate of the person's intelligence as indicated by IQ scores, it is also important to provide a discussion of more specific abilities. This discussion may include an analysis of areas such as memory, problem solving, abstract reasoning, concentration, and fund of information. If the report will be read by persons who are familiar with test theory, it may be sufficient to include IQ scores without an explanation of their normative significance. In most reports, it is helpful to include the IQ scores as well as the percentile ranking (see Appendix B) and general intellectual classification (high average, superior, etc.; see Table 5.1). Some examiners may even prefer to omit the actual IQ scores in favor of including only percentile rank and general classification. This can be useful in cases in which persons reading the report might be likely to misunderstand or misinterpret unexplained IQ scores. After a general estimate of intelligence has been made, it should, whenever possible, be followed by a discussion of the client's intellectual strengths and weaknesses. This may involve elaborating on the meaning of the difference between Verbal IQ and Performance IQ or a discussion of subtest scatter. In addition, it can be useful to compare the client's potential level of functioning with his or her actual performance. If there is a wide discrepancy between these two, reasons for this discrepancy should be offered. For example, the client may be underachieving because of anxiety, low motivation, emotional interference, or perceptual processing difficulties. Practitioners may also wish to discuss additional noncognitive areas of intellectual assessment. This might include the extent to which the person prefers to achieve through independent activities versus a structured environment, the level of motivation, or, the relative intellectual efficiency or hardiness. Cognitive assessments in psychiatric contexts might include any bizarre associations,

degree to which the person's thinking is organized, or how concrete or abstract his or her thought processes are.

Whereas a discussion of intellectual abilities is relatively clear and straightforward, the next sections are frequently more difficult to select. There are an extremely wide number of possibilities to choose from, many of which are listed in Table 15.1. Some practitioners recommend including set topics. These typically include the client's level of cognitive functioning, emotional functioning (affect and mood), and interpersonal relationships. A neuropsychological evaluation might divide the impression and interpretation into areas such as memory, language functions, executive abilities, awareness of deficits, sensory/perceptual functions, and personality (Groth-Marnat, 2000a; Hebben & Milberg, 2002). One rationale for not having a preset list of topics to discuss is that the topics should be based primarily on the referral question. This allows the practitioner to flexibly organize the topics based on the context of the report and the needs of the referral source and client. If the referral question is clearly focused on a specific problem, it may be necessary to elaborate on only two or three topics. A referral question that is more general may require a wider approach in which six or more areas are discussed.

Some additional common and important topics are the client's level of psychopathology, dependency, hostility, sexuality, interpersonal relationships, diagnosis, and behavioral predictions. A client's level of psychopathology refers to the relative severity of the disturbances he or she is experiencing. It is important to distinguish whether the results are characteristic of normals, outpatients, or inpatients, and whether the difficulties are long term or a reaction to current life stresses. Does the client use behaviors that are adaptive or those that are maladaptive and self-defeating? Within the area of ideation, are there persistent thoughts, delusions, hallucinations, loose associations, blocking of ideas, perseveration, or illogical thoughts? It may also be important to assess the adequacy of the client's judgments and relative degree of insight. Can the person effectively make plans, understand the impact he or she has on others, and judge the appropriateness of his or her behavior? To assess the likelihood of successful therapy, it is especially important to assess the client's level of insight. This includes assessing the person's ability to think psychologically, awareness of his or her own changing feelings, understanding of the behaviors of others, and ability to conceptualize and discuss relevant insights.

Usually, a client's greatest conflicts center on difficulties with dependency, hostility, and sexuality. In discussing a client's dependency, it is important to discuss the strength of these needs, the typical roles played with others, and present or past significant relationships. In what ways does the client defend himself or herself against, or cope with, feelings of dependency? This evaluation may include a discussion of defense mechanisms, thoughts, behaviors, feelings, or somatic responses as they relate to dependency. The relative intensity of a client's hostility is also important. Is the expression of hostility indirect, or is it direct in the form of either verbal criticisms or actual assaultive behavior? If the expression of hostility is covert, it may be the result of factors such as fear of loss of love, retaliation, or guilt. When the client does feel anger, what are his or her characteristic defenses against these feelings? For example, some clients might express opposite behaviors, with overly exaggerated concern for others, or they might direct their anger inward by developing physical aches and pains that serve as self-punishment for having aggressive impulses. They may also adapt through means such as extreme suspiciousness of others, which has been created by denying their hostility and attributing it to others. A discussion of a client's sexuality usually involves

noting the relative intensity of his or her urges and the degree of anxiety associated with the expression of those urges. Does the client inhibit his or her sexuality because of a belief that it is dirty, experience anxiety over possible consequences, or associate it with aggressiveness? Defenses against sexual urges may be handled in ways similar to hostility, such as by performing the opposite behavior through extreme religiosity or celibacy, or by denying the feelings and attributing them instead to others. On the other hand, clients may impulsively act out their sexual urges, at least in part, out of a need to obtain self-affirmation through sexual contact. Clinicians may want to discuss the dynamics involved in any unusual sexual practices.

Discussing clients' characteristic patterns and roles in interpersonal relationships can also be extremely useful. These can often be discussed in relationship to the dimensions of submissiveness/dominance and love/hate, or the extent to which they orient themselves around the need to be included, control others, or seek affection. Is their style of communicating typically guarded, or is it open and self-disclosing to the extent that they can discuss areas such as painful feelings and fears? Can they deal with the specifics of a situation, or are they usually vague and general? Do they usually appear assertive and direct, or passive and indirect? Finally, it is often important to determine the extent to which they are perceptive about interpersonal relationships and their typical approaches toward resolving conflict.

It may also be appropriate to include descriptions of vocational goals and aptitudes. This is becoming increasingly important in educational reports, especially for students with special educational needs, such as those with disabilities (Schalock et al., 1994). Many of the tests covered in this text can help in assessing a person's strengths and weaknesses, but practitioners may also need to include further assessment devices, such as the Self-Directed Search, Strong-Campbell Interest Inventory, or Kuder Occupational Interest Survey (see Prince & Heisser, 2000).

A frequent consideration is whether the client's difficulties will continue or, if currently absent, recur. If the client's future prospects are poor, a statement of the rationale for this conclusion should be given. For example, if a clinician predicts that the response to treatment will be poor, he or she should explain that this is caused by factors such as a strong need to appear hypernormal, poor insight, and a high level of defensiveness. Likewise, favorable predictions should include a summary of the client's assets and resources, such as psychological mindedness, motivation to change, and social supports. If difficulties are likely to be encountered during the course of treatment, the nature and intensity of these difficulties should be discussed. The prediction of suicidal potential, assaultive behavior, child abuse, or criminal behavior is essential in certain types of reports. Often the tests themselves are not useful in predicting these behaviors. For example, one of the best ways of predicting suicidal potential is to evaluate the client's history, current environment, personal resources, and degree of suicide intent (Klespies & Dettmer, 2000; Stelmachers, 1995). However, research indicates that many predictions of behavior, such as dangerousness, are subject to error (Binder, 1999; Freedman, 2001; Megargee, 1995). This is especially true for long-term predictions. Clinicians should thus exercise appropriate caution in making predictions and not exceed the bounds of reasonable certainty.

Sometimes clinicians may wish to include a separate section on diagnosis. However, whether to include a *DSM-IV* (1994; or *ICD-10;* World Health Organization, 1990) diagnosis has been an area of some controversy. Some clinicians feel that labels should be

avoided because they may result in self-fulfilling prophecies, be overly reductionistic, and allow clients to avoid responsibility for their own behavior. Other objections to diagnosis stem from researchers who feel that many of the terms are not scientifically valid (Beutler & Malik, 2002; Rosenhan, 1973) and are not particularly useful in planning interventions (Beutler & Malik, 2002; Groth-Marnat, Roberts, & Beutler, 2001; Houts, 2002). If a clinician does decide to give a diagnosis, he or she must first have a clear operational knowledge of the diagnostic terms. He or she should also include the client's premorbid level of adjustment, and the severity and frequency of the disturbance. Instruments such as the Structured Clinical Interview for the *DSM-IV* (SCID; First, Spitzer, et al., 1996; First et al., 1997), Structured Interview for *DSM-IV* Personality (SIDP), or Anxiety Disorders Interview Schedule (T. Brown et al., 1994) might help to increase the reliability of diagnosis. It is also important to include the possible causes of the disorder. A discussion of causes should not be simplistic and one-dimensional but rather should appreciate the complexity of causative factors. Thus, causes may be described from the perspective of primary, predisposing, precipitating, and reinforcing factors. Clinicians may also discuss the relative significance of biological, psychological, and sociocultural variables.

Summary and Recommendations

The purpose of the summary subsection is to restate succinctly the primary findings and conclusions. This requires that the practitioner select only the most important issues and that he or she be careful not to overwhelm the reader with needless details. As emphasized previously, a useful strategy in the summary section is to provide brief bulleted/numbered answers to each of the referral questions. It should also be noted that some practitioners prefer to place the summary subsection at the end of the Impressions and Interpretations section. Either location is acceptable; the choice can be based on the clinician's personal preference and the needs of the report as suggested by the referral question(s) and background of the readers.

The ultimate practical purpose of the report is contained in the recommendations because they suggest what steps can be taken to solve problems. Such recommendations should be clear, practical, and obtainable, and should relate directly to the purpose of the report. The best reports are those that help the referral sources and/or the clients solve the problems they are facing (Armengol, Moes, et al., 2001; Finn et al., 2001; Ownby, 1997; Tallent, 1993). To achieve this report-writing goal, the clinician must clearly understand the problem, the best alternatives for remediation, and the resources available in the community. One practical implication is that writers can improve their reports by becoming as familiar as possible with the uses to which their reports will be applied. An effective report must answer the referral question and have decisional value. After these factors have been carefully considered, recommendations can be developed. Decisions related to recommendations occur on three different levels (Beutler, 1995). First, decisions need to be made related to the *setting or context* (outpatient, day hospital, halfway house, inpatient, new work environment, change in schools/classes). Second, consideration needs to be given to *developing a relationship* with the client (degree of resistance, level of insight, interpersonal style, empathy, etc.). Finally, decisions need to be made about *specific intervention procedures* (systematic desensitization, emotional support, vocational training, rehabilitation, special education, etc.).

Practitioners and researchers alike have become increasingly interested in tailoring the results of assessment toward optimal client interventions (see Chapter 14). For example, Beutler et al. (2000) provide strong empirical support that planning interventions around problem complexity, problem severity, motivational distress, coping styles, level of resistance (reactance), and social support can increase the effectiveness of treatment outcomes. Additional relevant variables include the stage of change (Prochaska & DiClemente, 1992), hypnotic responsiveness (Groth-Marnat, 1991), neuroticism, level of stress (Wickramasekera, 1995a), level of abstract versus concrete thinking, and ego strength. Outcome research for particular types of problems (anxiety, sexual dysfunction, etc.) frequently indicates that certain interventions are more effective than others. Especially relevant are the specific causes and expression of a disorder. For example, depression can result from and be maintained by a variety of factors including irrational cognitions, poor social skills, grief, stress, poor interpersonal relations, medical illnesses, substance abuse, medications, and faulty biological mechanisms. The relative contribution of each of these causes has important implications for optimally tailoring client interventions. By considering each of these areas, the practitioner can develop optimal specific, targeted recommendations.

One clear finding is that reports are typically rated most useful when their recommendations are highly specific rather than general (Armengol, 2001; Finn et al., 2001; Ownby, 1990, 1997; G. White, Nielsen, & Prus, 1984). Thus, a recommendation that states "The client should begin psychotherapy" is not as useful as a statement of the need for "individual therapy focusing on the following areas: increased assertiveness, relaxation techniques for reducing anxiety, and increased awareness of the self-defeating patterns he creates in relationships." Likewise, a recommendation for "special education" can be improved by expanding it to "special education two hours a day, emphasizing exercises in auditory sequencing and increasing immediate recall for verbally relevant information." However, caution should be exercised when providing specific recommendations in some contexts because some health professionals may feel that developing treatment recommendations is primarily their responsibility or perhaps should be made by the overall treatment team (Tallent, 1993). After the report, with its recommendations, has been submitted, continued contact should be made with the readers(s) to make sure the report has not been filed and forgotten. Even the best report is not functional unless the recommendations are practical, obtainable, and actually put into action.

SAMPLE REPORTS

The sample reports in this section are from the more common settings in which clinicians work and consult. The dimensions in which the reports vary are:

- Format.
- Referral question.
- Extent to which history rather than test data is emphasized.
- Types of tests used.
- Degree to which they include a variety of descriptions rather than being case-focused with a relatively limited range of topics.

In each setting, specific questions have been presented, along with decisions that must be made related to the client. The different reports illustrate how the clinician has integrated the test data, client's history, and behavioral observations to handle these questions. The reports were selected to illustrate a wide diversity in format, length, type of setting, referral question, and type of tests used.

The first report was developed for a psychiatric setting and was intended to be read by professional mental health personnel. For this reason, there is some use of technical language and a focus on developing a detailed, traditional *DSM-IV* diagnosis. What is noteworthy in the handling of the test data is that the bulk of the discussion relating to test interpretation revolves around projective test findings (Rorschach, TAT, projective drawing). Furthermore, much of the projective data is used in a qualitative, content-oriented manner. This was achieved by providing actual verbatim responses, which give a more colorful and rich portrayal of the client's thought processes than could be achieved through quantitative scores. For example, some of the TAT stories are written out to illustrate attitudes toward the client's parents and how the client perceives and attempts to cope with his inner sense of "evilness." This is a more test-oriented approach than is recommended, but the report has the advantage of allowing the readers to see why and how certain inferences were made. For example, relevant Rorschach indices and TAT stories have been included adjacent to interpretations based on this data. For this reason, it is an excellent teaching case.

The second report, written in a legal context, critiques a report done by another professional who had evaluated a client for possible malingering. This is an important teaching case because, in the process of evaluating the report done by another professional, there is information on how the assessment should have optimally been performed. As such, it provides guidelines for proper assessment procedures. Specific points are the importance of using norms similar to the client that is being evaluated, the use of converging sources of information, placing the results into a wider context, and following acknowledged protocols for assessing client domains (in this case, malingering). The professional in question had used inappropriate norms (norms were for brain-injured patients complaining of memory difficulties and not pain patients), had used only one source of information (only one test of malingering), and had not considered the wider context (client characteristics, medical records). One feature to pay particular attention to is the clear listing of referral questions at the beginning of the report. This is followed in the body of the report by numbered paragraphs that respond to each of the referral questions. The main points are then summarized in bulleted form at the end of the report. Another aspect of the report is its format—written as an informal letter rather than according to structured headings (i.e., referral question, evaluation procedures, etc.).

The third sample is from an educational context; the client (a 12-year-old Hispanic female) was experiencing emotional and behavioral problems that were impacting her academic performance and peer relations. This report is an important addition because it not only includes a child, but also demonstrates how the examiner handled ethnic diversity. There is also information on how the combination of emotional difficulties and being bilingual was likely to have lowered her optimal level of cognitive performance. Two noteworthy strengths of the report were the extent to which a wide variety of sources of information were used (previous records, interviews with

several different persons, a number of formal tests) and the clear description of the client's assets. The formal tests were targeted toward specific areas (intelligence, achievement, trauma, self-perception) and used a combination of objective and projective/qualitative information.

The final report was written in a general psychology clinic and was intended for use by mental health professionals. As a result, some technical language is used, primarily in the form of a formal *DSM-IV* diagnosis. The major feature of the report is the extensive development of a detailed treatment plan for psychotherapeutic intervention. This plan was developed based on the Systematic Selection Model detailed in Chapter 14. The recommendations are eclectic in orientation and assume that the treating practitioner can effectively use a number of techniques from a variety of theoretical orientations. Another feature of the report is the absence of psychological test data. Specifically, there is no Test Results section because some, if not many, practitioners believe that the inclusion of detailed test results is both unnecessary and results in cluttering up the report with distracting detail. It is rather assumed that the referral source is most interested in the integration of the overall assessment along with the relevant recommendations.

THE PSYCHIATRIC SETTING*

NAME: Robert
DATE OF BIRTH: 2/5/86
DATES OF EVALUATION: 3/14/2000

REFERRAL QUESTIONS
Robert M. is a 14-year-old high school student admitted to the psychiatric unit at Monte Hospital on March 10, 2000, for his second psychiatric hospitalization. He was referred for an interview and psychological testing by Harold Smith, MD, to estimate his intelligence, differential diagnosis, behavioral dynamics, and potential for adjustment.

EVALUATION PROCEDURES
Robert was interviewed for one hour and administered the Wechsler Intelligence Scale for Children-3rd ed. (WISC-III), Rorschach, Thematic Apperception Test (TAT), Bender Visual Motor Gestalt Test (Bender) and Draw-A-Family Test. Information was also derived from a review of his hospital charts and a conversation with his psychiatrist (15 minutes; March 9, 2001).

BEHAVIORAL OBSERVATIONS
Robert was seen on the Delta unit at Monte Hospital on March 12, 2000. He presented himself looking his stated age and adequately nourished as a pubescent male with somewhat disheveled dark blond hair parted approximately on the right side, a fair facial complexion with mild acne and dressed in a white T-shirt, a blue and white plaid long-sleeved shirt in poor repair and new-appearing Levi's. A cross-shaped earring

* Report contributed by Tom MacSpeiden, PhD.

dangled from his left ear. At the beginning of the interview and multiple times thereafter, Robert asked when the evaluation would end. At one point, this examiner asked what Robert would do after the evaluation was completed, and Robert said he would go to his room to lie down because he was tired. He had a depressed facial expression, but was physically agitated and demonstrated considerable motor overflow although he remained continuously seated in his chair. During the intelligence testing, he seemed to exert effort until a task became difficult whereupon he became disheartened and reduced his effort. During the projective testing, he was brief and hurried to complete the tasks. There were no indications of delusions or hallucinations, and he was oriented in all three spheres. His hygiene was adequate. He was alert, and both recent and remote memory appeared intact. Given the previous behavioral observations and additional test indicators, I believe the assessment is an accurate assessment of his current level of functioning.

PRESENTING PROBLEM

Robert stated that he was in the hospital "for family problems." (What kind?) "Just not getting along. Not being able to work out at home." He said these problems started two years ago when "my mom put me into a private school, and I got kicked out" for "doing wrong things. Inappropriateness." Reportedly, on March 9, 2000, Robert consumed a large quantity of alcohol, became acutely intoxicated, and was agitated, belligerent, combative, tearful, and yelled obscenities. His family took him to Monte hospital where tests revealed a blood alcohol level of .24 without indication of other drugs. Seemingly, he consumed this near-lethal dose of alcohol in response to his hatred and fear of his parents.

BACKGROUND INFORMATION

The following information was obtained from Robert, who was a restless and dubious historian, from his hospital chart, and from a March 12, 2000, telephone conversation with Dr. Smith.

Robert was born and raised in San Francisco County and his biological parents remain married. The father, approximately age 40, is a career Air Force person with a rank unknown to Robert. He said his father teaches machine maintenance and attends college. He described his father as "a mean, violent, cruel person." (How does he show this?) "He used to hit me." (What would cause him to do that?) "Sometimes when I did something wrong." (Like what?) "Anything, any little thing, sometimes." He said he was hit with objects such as "a belt, sticks, spoons, forks, knives. Any object a lot of times." He said the most significant injuries he suffered during such beatings were "large blood blisters on my hand." He said he never suffered an injury that required medical attention but that he did ask his father's permission to remain out of school to conceal the marks from his peers. He said his father refused the request.

Robert's mother is approximately age 35 and a business secretary. Asked to describe her, he said, "I don't know. I guess she's mean. That's what I think. Is it my opinion?" (Yes. Mean in what way?) "Verbal abuse, yelling, screaming." (What causes her to do this?) "Anything, like my father. Frustration." (How does she react when your father hits you?) "She doesn't do anything. She lets him."

Robert is the oldest of three siblings with a brother, age 13, and a sister, age 10. Both siblings live in the home and attend school. "My brother's a little strange. Very silly and inappropriate for his age." He said his brother shows these characteristics when he "runs around naked, talks real strange like he's on an acid trip when he's not." He said his brother excels in school and has never received counseling or psychotherapy. Regarding the sister, "She's fine."

Medically, Robert believes he experienced an uneventful gestation and delivery. Regarding unusual illnesses or accidents, he said that at age 10 or 11 he broke either his wrist or his thumb during horseplay with friends and that the break healed without limitation. He said that at a later age he could not recall he broke his finger and wrist in horseplay and that he had broken such bones approximately four or five times. He said that otherwise his health has been good.

Asked about unusual events in his life before starting school, Robert said, "I can't remember that far back." He said he enrolled at Rutledge Elementary School at the usual age where he remained in regular classes through the sixth grade. He earned primarily "C's," and related to his peers "well." He said he also related to his teachers "well." He said he had no behavioral problems in elementary school and that his life was uneventful.

For the seventh grade, Robert was placed in a Roman Catholic school "because during the sixth grade they felt I was failing, not doing my work, taking too much leisure time." He said he remained at the Catholic school until the last month of the school year when he was expelled because of frequent misbehavior. Initially, he said he could not remember what act led to the expulsion but later said, "I lit a firecracker in class." (In your opinion, why did you do that?) "Rebelling." (Against?) "Parents. It was also for fun." (Why were you rebelling against them?) "Cause I didn't want to go to that school. I wanted to go to public school." (Why?) "Because I had gone to a public school for six years, and all my friends were going there."

Robert said he remained home without attending school until the following fall, September 1999, when he enrolled in the eighth grade at Franklin Junior High School to remain in regular classes until entering Monte Hospital. In the eighth grade, his marks were "failing, no, below standard." (Why were they so low?) "Because I had started using alcohol, drugs. I don't know. I don't know if that's the reason." (Why did you start using drugs?) "I don't know that, the reason why." He said he was introduced to alcohol and other drugs by his age peers. He estimated that he was intoxicated approximately 30 times during the eighth grade, usually at home with friends while his parents were out. The only street drug used was marijuana. He said his parents did not know about his substance misuse but complained about his grades. "They wanted me to bring them up, but I started failing even more." He said otherwise there were no unusual events during the eighth grade.

Last September, Robert enrolled in the ninth grade. Then, "I started hanging out with more friends that drank and used more drugs," but he misused only alcohol and marijuana. He said he was intoxicated "a few times a week" and misused marijuana "maybe four or five times a week." He said that otherwise his life was uneventful until he entered Monte Hospital for the first time during Christmas vacation because "I was dead drunk." He said he remained hospitalized three days, until "the doctor thought I

should be discharged." A week later, he returned to Monte Hospital for his current hospitalization.

Asked about his future plans in life, Robert said, "I haven't made any. I want to live with my grandparents, but they're not going to let me." He said his mother and his treating psychiatrist were against this plan because "they think I'm running away from my problems."

INTERPRETATIONS AND IMPRESSIONS

On the WISC-III, Robert obtained a Verbal IQ of 91, a Performance IQ of 84 and a Full Scale IQ of 86. His Verbal scaled scores ranged from 7 to 10 without significant deviation from the Verbal mean. Similarly, his Performance scaled scores ranged from 6 to 10. Verbally, he functioned at the 27th percentile and in the Average range. In terms of perceptual motor performance, he functioned at the 14th percentile and in the Low Average range. Overall, he functioned intellectually at the 18th percentile and in the Low Average range. His performance is probably a conservative reflection of his intellectual capacity, which may be as high as the high average range.

The structure of Robert's personality was difficult to estimate from his limited 14 Rorschach responses but appears nonpsychotic. Because of his limited responses, his ratios and percentages should be interpreted with caution. He uses repression and denial excessively in an attempt to exclude unacceptable impulses and fantasies from consciousness ($F = 92\%$). Nonetheless, Robert has difficulty seeing those things in the environment seen by most persons (Populars = 1) and at times gravely distorts his perceptions ($X + \% = 65\%$). These distortions are not frequent and are most likely to occur when he is emotionally stimulated. Because of his rigid defensive posture, he currently has few energies available ($M + C = 2.0$). He attempts to enhance control by denying emotional stimuli ($Sum\ C = 0$). And generally when expressing emotion, he has adequate control ($FC : CF + C = 1 : 0$). Because of his rigidity, his aspirations exceed his ability to perform ($W : M = 7 : 0$), and there is no indication of whether he prefers to accomplish more in fantasy or in action ($EB + 1 : 1$). Because of his excessive repression and denial, his depression is not obvious as he does not easily respond to emotional material ($Afr = .60$). For his age, his interests are somewhat narrow and immature ($A = 64\%$).

The content of Robert's personality projected in his Rorschach responses is a perception of himself as confused and disorganized, much as he projected in his first response to Card I: "A star; a messed up star. Someone drew it wrong." As he noted in his Rorschach scores, he attempts to deny emotional stimuli to enhance control, a mechanism typified in his rejection of Cards VII and IX, the first of three highly chromatic and, therefore, affectively stimulating cards. When presented with Card VIII, he responded in approximately three seconds, "I see nothing in this, I don't see anything." His response to Card IX was similar. On Card X he used the whole of the blot as an "insect," a response that had negative form level. Cautiously, he saw on each of the first five Rorschach cards the percept of "insect." Such perseveration is often found among neurologically damaged persons although it appears more likely in Robert's case a product of anxiety that limited his ability to structure the amorphous blot material much beyond an unspecified insect.

The content of Robert's personality as projected in his TAT responses is a painful perception of himself as evil, a perception that may have resulted in part from his already overly punitive superego receiving further stimulation while he was enrolled in a parochial school for the seventh grade. The content suggested by his cryptic response to Card 2 (man plowing with a horse in the background, a pregnant woman standing to one side near a tree, and a girl in the foreground holding books) was:

> Past, going, gone to a church school. Present, leaving home. Future, death. (Death in what way?) Starvation.

Robert views his internal anger and violence as undeniable indicators of his evilness, and evilness from which he can escape by drowning it along with himself in alcohol. Vivid were his responses to Cards 8BM (a primitive surgical scene with a boy standing in the foreground and a gun to one side) and 13MF (a bare-breasted woman lying on a bed and a man standing with one hand to his forehead):

> 8BM Past, Vietnam War. Present, remembering. Future, death from a heart attack. (What was he remembering?) The people being cut up in the Vietnam War.
>
> 13MF Past, alcoholic. Present, wife dead. Future, death from alcohol. (Why did the wife die?) Kidney disease.

Alternatively, he conceptualizes continuing to live with the evil in him without death as projected in his response to Card 15, a figure standing among tombstones:

> Past, possessed with evil. Present, trying to destroy the evil. Future, living with evil. (What was the evil?) The devil.

He believes that were he to continue living with his evil, he would represent a danger to other persons. This content was typified in his response to Card 7BM, a younger and older man in conversation:

> Past, jail. Present, courtroom. Future, jail. (Why was he in jail?) Murder.

Robert believes that in the past he was unwanted by a significant other. This awareness instilled in him the belief he was lacking in some way. When someone now offers him acceptance, he rejects it although he hopes he will be able to accept it at some point in the future.

On the Bender, all figures were completed with relative accuracy, and there were no indications of a neurological dysfunction affecting Robert's perceptual motor control. There were several indicators of impulsivity including overlapping, poor planning, and poor overall quality.

Asked to draw a picture of his family, Robert drew from left to right persons he later labeled verbally as "Dad," "Sister," and "Mom." Above these three and along the upper

margin of the page from left to right, he drew "Me" and "Brother." The drawing projects Robert's perception of him and his brother as separate from the parents and sister, much as he indicated during the interview when he said, "My brother's a little strange"; and clearly, he views himself as somewhat strange. Not surprisingly, he perceives the father as powerful and his mother as significantly weaker than his father.

DIAGNOSTIC IMPRESSIONS *DSM-IV*

AXIS I	300.40	Dysthymia, Secondary Type, Early Onset
	303.90	Alcohol Dependence, 305.20 Moderate Cannabis Abuse
AXIS II	v71.09	No Developmental or Personality Disorder Diagnosed
AXIS III		No Physical Condition or Disorder Diagnosed
AXIS IV		Psychosocial Stressors: Perception of himself as intrinsically evil and unlovable, believed rejection by parents, effects of drug misuse, failure to adjust socially and psychiatric hospitalization. Severity: 4-Severe (Admixture of acute events and enduring circumstances)
AXIS V		Global Assessment of Functioning (GAF): Current GAF: 35 Highest GAF past year: 50

SUMMARY AND RECOMMENDATIONS

The previous information suggests the following answers to the referral questions:

- Robert functions intellectually in the low average range although his endowment may be in the high average range and is currently reduced by his mental condition. He does not suffer from a neurological dysfunction affecting his perceptual motor control.
- His personality structure is nonpsychotic and his difficulties are more consistent with dystyhmia with alcohol dependence and moderate cannabis use.
- His personality structure is exceedingly rigid and characterized by repression and denial such that he fails to see those things in the environment seen by most persons and has few energies available. He attempts to enhance control by denying emotional stimuli and generally can maintain control when responding emotionally. When highly stimulated emotionally, however, he gravely distorts his perceptions. With his rigid defense structure, he has few energies available to gratify his aspirations but is capable of concealing overt indications of depression such as tearfulness and significant withdrawal. When he is consuming disinhibiting chemicals such as alcohol, his behavioral and emotional expressions are far less controlled.
- Robert has at least a vague awareness he is "messed up." It is probable he was highly critical of himself before enrollment in the Roman Catholic school. There he may have been overly scrupulous when interpreting moral teachings, and his already punitive superego became self-destructive. He may have felt most guilty

about the commandment to honor parental figures. He came to view himself as anathema. He believes his alternatives are to continue living with the believed devil inside him with the probability he will eventually murder someone or, alternatively, to end his life to escape the internal evil and avoid harming others. Not unexpectedly, death is frightening to him and presently he flirts with self-destruction by poisoning himself with alcohol.

A more complete social history would be useful in determining the major factors contributing to Robert's strong negative self-image. Although it is unlikely Robert's father is as abusive as Robert describes him, some series of events have led Robert to view his father as unaccepting and unacceptable. A combination of family psychotherapy on a weekly basis and continuing individual psychotherapy with Robert are the treatments of choice. Initially, Robert will find individual treatment threatening and, consequently, may gain more in the family sessions. As he progresses in treatment, he will profit more from the individual sessions. Without treatment, he is a significant danger to himself and could end his life in suicide.

THE LEGAL CONTEXT

Dear Mr. Hamlin:

Thank you for the opportunity of commenting on the March 21, 2000 report of James A. Speculus, MD, undertaken as an evaluation of Mrs. Sharon Paine. As you know, Mrs. Paine is a 35-year-old, Caucasian married female with a university education who sustained a motor vehicle accident on May 17, 1998. She experienced injuries to her mouth and left shoulder that have caused her ongoing physical pain, several surgeries, and emotional distress. In addition, she has evidently had difficulties adjusting to her condition, which has precipitated symptoms of insomnia, depression, feelings of hopelessness, irritability, anxiety, and physical pain. According to the March 21, 2000 report of Dr. Speculus, these symptoms had mostly been resolved at the time of his assessment and he concluded that the complaints were mostly due to malingering. It is my understanding that you would like me to comment on the methodology in Dr. Speculus' report with specific questions related to the validity of the Test of Memory Malingering (TOMM), the appropriateness of using the TOMM with the client, an optimal protocol for assessing malingering for a client such as Mrs. Paine, possible reasons why Mrs. Paine would have scored poorly on the TOMM, and the training/qualifications required for a person using a test such as the TOMM. I note in this regard that I was neither requested to actually evaluate the client's condition, nor did I actually meet with her directly. Rather, I will contain my comments to the methodology and instruments used by Dr. Speculus.

For clarity and ease of reference, my comments are numbered:

1. The TOMM is essentially a well-designed, psychometrically sound test designed for assessing the exaggeration of memory deficits among persons with known or suspected neuropsychological impairment. By this, I mean that it is fairly accurate in

distinguishing persons simulating memory deficits from those who are normal and even those who have well-documented, physically based memory deficits (Rees, Tombaugh, Gansler, & Moczynski, 1998; Tombaugh, 1997). Thus, it is a good test for the purposes for which it was designed. However, one of the weaknesses of the TOMM, as stated by researchers, is a "failure to include a psychiatric control group to evaluate the possible effects of psychiatric symptomology, such as depression, on the TOMM" (Rees et al., 1998, p. 18).

2. In evaluating the appropriateness of the TOMM for Mrs. Paine, it is essential to point out that Mrs. Paine was presenting with and being evaluated for symptoms of pain, depression, and anxiety. Thus, it is confusing why a test designed to assess the possibility of exaggerating *memory* deficits was used. Indeed, I note that there is an extensive literature and a number of well-developed tests that have been specifically designed to detect the possible exaggeration of emotional (depression and anxiety) difficulties. Some of these tests have the advantages of not only assessing for the validity of symptoms, but can also assess a wide range of highly relevant emotional and personality variables. Thus, I would have thought that Dr. Speculus would have tailored his test selection based on the presenting problem of the client. I conclude then, that although the TOMM is indeed a good test for the purposes for which it was designed, in this case it was not being used for the purpose for which it was designed.

3. A further issue is that a single test, particularly one such as the TOMM, should never be used in and of itself to draw final conclusions about a client. Standard texts on assessment all stress the importance of integrating multiple sources of information in making final conclusions (see Beutler & Berren, 1995; Groth-Marnat, 1999, 2000a). In contrast, the format of Dr. Speculus' March 21, 2000 report suggests that he used a mechanical, unintegrated, test-dominated manner of making his conclusions ("I asked her to perform the TOMM. She scored 35 out of 50 . . . this indicates an intention to deceive me." p. 6). This issue of integrating converging sources of information is particularly crucial in the case of malingering where a conclusion that the client has malingered can have a major impact on the client's life. A protocol for assessing malingering as recommended by major publications in the field (see Blau, 1998; Franzen & Iverson, 1998; R. Rogers, 1997) would include not only information on a single (appropriately used) test and the context of possible gain, but also relevant dimensions such as personality (i.e., antisocial trends), attitude (i.e., "the world owes me a living"), past history suggestive of using people/organizations for gain, collateral interviews with outside sources, multiple observations of the client's behavior, and patterns among several test results (i.e., inconsistent patterns between tests). My reading of the report indicates that few of these dimensions were evaluated. (Although there was a comment that "she was careful about choosing her words," care in wording would be expected for the majority of people being evaluated, particularly a university-educated person such as Mrs. Paine.).

4. Despite the inappropriateness of using the TOMM and its apparent lack of integration with additional relevant (mostly unassessed) aspects of the client, it still remains unclear as to why she scored as low as she did. Possibilities include an undiagnosed neuropsychological condition, exaggeration of resolved (but currently minimal) difficulties, exaggeration of actual quite significant difficulties, or the impact of emotional factors on test performance. Given her age and history, an undiagnosed neuropsychological condition such as dementia is unlikely. A further option (taken by

Dr. Speculus) is that she was exaggerating minimal (but now "resolved"; see p. 6 of report) difficulties. In contrast, another option might be that there was some exaggeration, but there were also quite significant difficulties the client was experiencing. Indeed, the context and process of medico-legal evaluation sometimes encourages clients with actual difficulties to make sure their difficulties are noticed by the client exaggerating these difficulties. This can clearly complicate the process of evaluation and make assessment of the true extent of difficulties quite challenging. I note in this regard that Dr. Speculus' report did not seem to consider any of these options. I would speculate that, given the recent death of Mrs. Paine's father (4 to 5 weeks before the March 21, 2000 report), shoulder injuries, and ongoing palate pathology, it would be unusual if Mrs. Paine were not experiencing a relatively high amount of distress. Finally, as indicated in the Rees et al. (1998) study, cognitive tests can be lowered by emotional factors (e.g., Burt et al., 1995; Finlayson & Bird, 1991; Sbordone, 2000a). These last two possibilities (exaggeration but with actual significant underlying difficulties and lowering of test performance due to emotional factors) seem to have been insufficiently explored as possible reasons for the lowered TOMM performance.

5. Training in the appropriate use of psychometric instruments involves a relevant professional degree, minimum of two full semester units in psychological testing, a semester on statistics, additional relevant course work (personality theory, psychopathology, psychotherapy, etc.), and two years of supervised experiences of which a substantial portion should relate to the use of tests with actual clients. Key features of this training would emphasize the importance of evaluating a test's norms in relation to whom the test will be used on, tailoring tests for the specific presenting problem, developing an optimal assessment protocol based on relevant research, and integrating test scores with a wide range of relevant information (see Groth-Marnat, 1999). I note that test developers do screen potential purchasers of restricted tests (such as the TOMM) by using a brief series of questions. However, these brief screening questions do not guarantee that this minimal training has actually been undertaken.

In summary, the assessment of Mrs. Paine did seem to elicit a number of seemingly relevant aspects of her history. However, the assessment of possible malingering was not optimal because:

- The test selected (TOMM) was not tailored toward her presenting problem.
- The test was not normed for clients similar to Mrs. Paine.
- The interpretation of the TOMM appeared to be done in a mechanical, test-dominated manner.
- Alternative possibilities for the meaning of Mrs. Paine's score on the TOMM did not appear to be sufficiently explored.
- The assessment of malingering appeared to include only a minimum number of relevant variables.

I hope that the previous perspective and comments help to better evaluate Mrs. Paine's case. If you have any questions or would like further elaboration, please contact me at your convenience.

THE EDUCATIONAL SETTING*

NAME: Anna S.
DOB: 05/21/89
GENDER: Female
DATES OF ASSESSMENT: July 17, 21, 22, 24, 2001
DATE OF REPORT: August 4, 2001
EXAMINER: Annie Chung, Ph.D.

REFERRAL QUESTION

Anna is a 12-year, 1-month-old bilingual, Hispanic female who is currently residing in a foster home due to parental abuse, neglect, and abandonment. Anna presents with irritability, anxiety, and poor peer relations. Additionally, Anna's foster parents indicate that Anna is having problems in school because of interpersonal difficulties with her classmates and problems adjusting to their home (e.g., seeks constant attention, argues with foster siblings). Anna's therapist, Mary Smith, LCSW, requested this psychological evaluation to assist in assessing her current level of cognitive and emotional functioning, clarify her diagnosis, identify her strengths, and provide suggestions that may be useful in helping her to adjust to school.

EVALUATION PROCEDURES/SOURCES OF INFORMATION

Clinical interview with Anna.

Review of social service/family court documents.

Review of current case file.

Consultation with Anna's social worker, therapist, teacher, and foster parents.

Wechsler Intelligence Scale for Children-III (WISC-III).

Wechsler Individual Achievement Test—Screener (WIAT).

Trauma Symptom Checklist for Children (TSCC).

Personality Inventory for Youth (PIY).

Children's Apperceptive Test.

Draw-A-Person/House-Tree-Person.

Child Self-Report and Projective Inventory (Color How You Feel, Critical Items, Color How Others Make You Feel, Perceived Competence, Draw a Child in the Rain, Sentence Completion, Kinetic Family Drawing).

BEHAVIORAL OBSERVATIONS/MENTAL HEALTH STATUS

During the initial testing session, Anna presented as a pleasant and engaging youth who was casually dressed and appropriately groomed. She appeared to be of average size and weight for her gender and chronological age. She willingly accompanied this examiner and appeared eager for individualized attention. Anna remained cooperative, maintained good eye contact, and readily responded to questions with descriptive

* Report submitted by Annie Chung, PhD.

details. She was clearly oriented to person, place, time, and situation. Her mood was somewhat depressed and her affect was generally constricted. She indicated that the reason for this evaluation "is to see how I'm doing." Anna demonstrated fair insight regarding her current situation, identifying her mother's abandonment of her as the reason for her current foster care placement and her associated feelings of distress. Her responses to hypothetical situations requiring decision-making skills suggest somewhat compromised judgment. Thought content was appropriate to the situation and thought processes were lucid, concrete, and coherent. There was no evidence of perceptual disturbances, flight of ideas, circumstantiality, or loose associations. Cognitive abilities appeared to be within the average range. Although not formally assessed, her conversational English skills were apparently proficient. When responding to direct questions, Anna typically responded in grammatically correct English. At times, she responded in Spanish but quickly restated her response in English, switching between both languages with apparent ease. During subsequent testing sessions, Anna's test-taking behaviors continued to be cooperative. Her approach to each task was marked by diligence, good effort, eagerness to please the examiner, and heightened performance anxiety (e.g., "Am I doing a good job?"). Although Anna exhibited good attention and concentration with the majority of testing activities, she also worked rapidly when given written tasks. Anna reported that she is "good at writing fast" and stated that she wanted this examiner to see how quickly she can work.

Even though Anna appeared to give her best efforts to the tasks presented to her, she is a bilingual youth whose first language is Spanish. Given that the tests administered were not specifically normed on this population, the assessment results should be treated with caution. In addition, she may have somewhat overreported her current level of emotional difficulties.

BACKGROUND INFORMATION

Information obtained from family court documents, consultation with Anna's therapist, her social worker, her current foster parents, and interview with Anna indicates that Anna has experienced multiple disruptions in her home life events throughout her childhood. These events include physical/emotional abuse, neglect, and abandonment. Anna lived with her biological mother, Lupe S., until she was approximately 6 years old. At that time, Anna and Ms. S. moved in with Juan F., whom Anna refers to as her stepfather. Anna continued to reside primarily with her mother, stepfather, and their two children, Roberto and Juanita, until March 2001.

After moving in with her stepfather, Anna reported that her mother repeatedly stated, "I wish you [Anna] were never born," and that her parents frequently yelled at her, hit her, and did not provide adequate clothing or care for her. Furthermore, Anna reported that she and her mother frequently left her stepfather's home to live with other relatives or in homeless shelters because of her stepfather's "mean" behaviors but that her mother always returned to him. Family court documents indicate that there were five substantiated reports of physical/emotional abuse and neglect of Anna and/or her siblings by both her mother and stepfather. Anna's parents reportedly did not comply with family supervision plans as mandated by Child Protective Services. Additionally, Anna explained that over the past several years her mother exhibited "weird" behaviors such as rummaging through garbage cans, forgetting to wash clothes, and most pertinent

to the child, witnessing her mother's attempt to choke herself. Anna also reported that her mother has been hospitalized at least 4 to 5 times for attempting to kill herself. In the past year, Anna was temporarily placed at the local children's shelter after running away from her home and when her stepfather attempted to abandon her at this facility. Subsequently, Anna resided with a maternal aunt for an undetermined amount of time before returning to her mother's care in the spring of this year.

During the spring of 2001, Anna and Ms. S. resided intermittently with relatives, Ms. S.'s employer, and then in a homeless shelter. Ms. S. then requested that acquaintances from church care for Anna. The acquaintances agreed to this on a temporary basis. However, they reported that Anna began having significant problems at school (e.g., suspended for fighting). Furthermore, these caretakers stated that Anna appeared to have significant "emotional problems" and were, therefore, unwilling to care for her any longer. They contacted Social Services and explained Anna's situation, clarifying that her mother was not available to care for her. As a result, Anna has been a ward of the state since May 2001, and has been placed in the foster home of Mr. and Mrs. G.

Anna's social worker referred Anna for mental health services given her history of neglect and abuse, physical aggression in school, and her excessive irritability. On placement in her current foster home, Anna initially appeared very sad and had difficulties interacting with her foster siblings. Anna's foster parents indicated that Anna has not exhibited overt behavioral problems but does appear to be immature (e.g., excessive teasing, difficulty sharing), needs a considerable amount of individual attention, and displays excessive moodiness.

INTERPRETATIONS AND IMPRESSIONS

Cognitive Functioning. On the WISC-III, Anna achieved a Full Scale IQ of 85; Verbal IQ of 87; and Performance IQ of 86. Overall, this places her in the Low Average range or 16th percentile (lower 16% of the population) when compared to her age-related peers.

A close examination of Anna's relative strengths indicates relatively intact visual-spatial/mechanical skills (Perceptual Organization Index = 91) as well as good verbal expression/conceptualization skills associated with school-related learning (Verbal Comprehension Index = 89). She is likely to do particularly well in tasks requiring her to understand social norms and follow the meaning of nonverbal social situations. This was consistent with this examiner's observations of Anna during the interview in which she responded appropriately and answered all questions clearly.

In contrast to these strengths, Anna's responses to tasks that required her to respond quickly and accurately to nonverbal information were quite low (Processing Speed = 67). This suggests that even though she can understand what might be expected of her in school, she will do poorly when she is required to perform under time constraints. In these conditions, she would be likely to sacrifice speed for accuracy. In addition, she may have a difficult time concentrating, especially when dealing with numerical information presented to her verbally (e.g., recalling phone numbers, making arithmetical calculations).

In comparison to her low average global cognitive potential (WISC-III FSIQ = 85), Anna's composite academic achievement skills (WIAT Screener; Total *SS* = 93) is

within the average range. Anna's scores on the Spelling (*SS* = 102) and Basic Reading (*SS* = 98) subtests suggest that these are relative strengths while her Mathematical Reasoning (*SS* = 81) skills are considerably weaker. However, Anna's solid average performance in Spelling and Basic Reading, which actually exceed her global cognitive potential, suggests that Anna's scores on the WISC-III may be an underestimate of her true cognitive potential. This is likely because of her bilingualism and the absence of normative data pertaining to this specific population's performance on standardized intellectual assessment measures. Her lower scores on the IQ tests are also likely to result from her difficulties working under time constraints.

Emotional Functioning. Anna's reports of multiple and atypically severe intrapersonal discomfort is associated with her traumatic and disruptive childhood experiences. She is experiencing heightened levels of internal anxiety, depression, and uncertainty as well as low self-esteem. For example, she frequently provided self-deprecating remarks about her drawings (e.g., "it's ugly"), and the colors she used to represent her feelings depicted her emotional states as overwhelmingly sad and worried. She seems to harbor feelings of anger but has difficulty expressing these appropriately and, thus, may have a propensity to act out. However, it also appears that she attempts to mask her internal distress by presenting herself as outwardly happy and content. Furthermore, Anna depicts her life experiences as extremely stressful and has significant difficulty coping with such experiences. Her perceptions of her family and others indicate significant ambivalence, isolation, and alienation. For example, when prompted to draw a picture of her family, Anna did not include herself and depicted all family members enjoying activities independent of each other. However, she also incorporated symbols of warmth and closeness in her house drawings. Thus, Anna appears to hold conflictual feelings about her family while she also hopes for safety and support.

In response to questions eliciting Anna's experience of critical life events, she identified three specific traumatic experiences that appear to be associated with her current worries, fears, nightmares, and sadness. Anna reported that when she was about 8 years old, a homeless man attempted to take her while she was walking by herself in the evening but that she never shared this experience with her parents because she felt that it would not be important to them. Additionally, Anna explained that when she was about 10 years old, she witnessed her mother attempting to choke herself and subsequently worries that her mother will kill herself. Most recently, Anna reported that her mother left her with people that she did not know and has been feeling very sad and worried since that time. Anna indicated that she has difficulty staying asleep throughout the night as she has bad dreams about these "scary" events. She also admitted to having considerable difficulty concentrating at school because of intrusive thoughts about her past experiences and worries about her future.

Anna's interpretation of story cards was representative of overwhelmingly negative outcomes for the child figure. Themes included betrayal by mother, excessive worries and fears, abandonment and rejection by others, lack of caring and attention from mother, foreshortened future, harm inflicted by others, and inability to obtain wishes and desires such as reunification with family. Thus, it appears that Anna experiences her world as generally unsafe, unpredictable, and lacking emotional warmth. However,

in one story, Anna projected a positive resolution for the child, in which the parents comforted and cared for the child "when she needed them." This suggests that Anna has a strong desire to be nurtured and accepted by caregivers.

In spite of these traumatic and stressful life experiences, Anna is able to identify positive self-attributes related to her academic skills (e.g., spelling, handwriting), personal appearance, and athleticism. It appears that these strengths and perceived competencies bolster Anna's self-esteem and promote her ability to function on a day-to-day basis and hope for positive changes in her life.

Interpersonal Functioning. Anna typically responds to people with the hopes of obtaining attention and acceptance. This has been evident in her interactions with adults, which have been characterized by a strong eagerness to please, requests for individual attention, and gift-giving. However, Anna's negative experiences with her parents have likely bolstered her defenses against continued rejection. Therefore, Anna may respond by rejecting others first. This has been exemplified in her history of aggressive peer interactions, which have led to persistent teasing and fighting at school as well as her own reports of not having any "true" friends. While such behaviors have not been problematic at her current placement, Anna's foster parents have reported that her social interactions with peers have been immature (e.g., whining, teasing, difficulty sharing). Yet, when provided with support and nurturing, Anna has become increasingly able to verbalize her difficulties and share painful experiences with adults (e.g., therapist, foster mother).

Behavioral Functioning. Assessment of Anna's overt behavior did not reveal high levels of acting out (e.g., aggression, delinquency). In contrast, she is likely to internalize her distress by becoming anxious and depressed. Anna's foster mother, Mrs. G., described Anna as initially very argumentative and oppositional but said that these behaviors subsided within the first several weeks in the new home. Overall, Mrs. G.'s ratings of Anna's current functioning suggest that Anna is not displaying significant behavioral concerns but that she does demonstrate symptoms of emotional distress at this time. It appears that Anna is responding positively to the structure, consistency, and nurturing provided by her foster parents although concerns associated with Anna's internal turmoil are evident.

DIAGNOSTIC IMPRESSIONS

AXIS I	309.81	Posttraumatic Stress Disorder
	995.5	Neglect/Physical Abuse of Child
AXIS II	v71.09	No Diagnosis on AXIS II
AXIS III		None known/reported
AXIS IV		Abandonment by mother
AXIS V		Current GAF: 60 Highest GAF past year: 50

SUMMARY AND RECOMMENDATIONS

Findings from this psychological evaluation indicate that: (a) Anna's academic skills in reading and spelling exceed her measured cognitive abilities, (b) she does not currently exhibit overt behavioral difficulties although peer/adult interactions have been strained, and (c) she is demonstrating and experiencing significant signs of emotional distress. While Anna's global cognitive ability score is within the low average range, it is likely an underestimate of her true ability level given her higher than expected scores on spelling and reading skills. This suggests that Anna's bilingualism and slowed performance may have inhibited her ability to demonstrate her optimal cognitive ability on the WISC-III. It is also quite possible that Anna's exposure to early and persistent abuse/neglect may have impeded the development of her optimal cognitive functioning, which is likely to be in the average range. Anna's experience of multiple traumatic events, including the threat of an abduction by a stranger, witnessing her mother's suicide attempts, and abandonment, have further exacerbated her feelings of fear/uncertainty associated with persistent maltreatment by her parents. It is evident that Anna's worries, sadness, difficulty concentrating, and recurring nightmares are associated with these traumatic experiences. Subsequently, Anna's internal turmoil has been increasingly evident in her interpersonal relationships. With adults, Anna exhibits a strong desire to please and seeks constant individual attention. With peers, her attempts to be recognized have included a large repertoire of negative attention-seeking behaviors. Although obstructive, these patterns of relating to others appear to represent Anna's attempts to obtain any type of personal acknowledgment, albeit negative, in a world that she perceives as largely unsafe and threatening.

However, it is also important to highlight Anna's resiliency in the face of immense risk factors in her young life. Anna's ability to acquire new knowledge (e.g., proficiency in English) and her attempts to initiate relationships with others suggest that she possesses the capability to succeed in academic settings and is receptive to building social support networks outside of her family of origin. In spite of her current and expected difficulties associated with trauma, Anna presents as a likeable youth who desires acceptance from others. Ensuring placement in a loving, consistent, and structured home environment that provides unconditional acceptance is absolutely crucial in supporting Anna's cognitive and socioemotional development and long-term stability. This is also likely to improve her cognitive level of functioning and enhance her academic performance.

Given the previous findings, the following recommendations seem appropriate:

1. Ensure regular attendance at school to bolster Anna's academic skills. Providing extra assistance/tutoring for mathematics and strategies on how best to work with timed tasks would be particularly helpful.
2. Provide Anna with opportunities to participate in structured community activities of interest to her (e.g., sports team) to promote her sense of self-efficacy and enable her to develop positive peer relationships in natural settings.
3. Provide Anna's foster parents with ongoing parenting education and parenting skills training to promote their understanding of Anna's needs (e.g., supportive structure, limit setting, explanations, consistency, acceptance) and use of

specific strategies to promote Anna's positive adjustment to her new placement, as well as school.

4. Reevaluate Anna's cognitive and academic functioning, including language proficiency in English and Spanish, after stable home and academic placements are secured, to determine if a need for additional educational interventions are warranted.
5. Provide Anna with individual therapy that uses supportive as well as cognitive-behavioral strategies to enhance her ability to further process her thoughts/feelings about her situation, promote her self-esteem, facilitate her use of positive coping strategies when distressed, and alleviate her trauma-related symptoms. As therapy progresses, it would be important to help Anna understand and process her experience of her family, her abandonment, and abuse. Integrating and resolving these issues will be essential in assisting her to become a well-adapted adolescent and young adult.

THE PSYCHOLOGY CLINIC*

NAME: A. G.
DATE OF BIRTH: May 30, 1925
DATE OF EXAMINATION: December 12, 1995
CASE #: 96–041
SEX: Female
ETHNICITY: European American
REFERRED BY: Dr. M.

REFERRAL QUESTION

This 70-year-old, divorced woman was referred for psychological evaluation by Dr. M., who specified that the patient suffered from agoraphobia and requested assistance in identifying effective treatment for her condition. A. G., on the other hand, indicated that she had had agoraphobia in the past, but described her current problem as one of motivation rather than panic or fear. She asserted that she is currently not immobilized nor is she extremely anxious when she travels. She attributes her restrictive lifestyle to the absence of "energy" and "motivation" to travel and engage in social activities. She acknowledged an "underlying apprehension" that arises when she is scheduled to leave home, however, resulting in her putting off her departure as long as possible. After she actually goes out, she reported that she remains anxious until she returns home.

The current evaluation was designed to clarify the nature of the problem, to develop treatment plans, and, if indicated, to initiate a treatment program.

EVALUATION PROCEDURES

By prior arrangement, A. G. completed the Life History Questionnaire before her intake interview on December 12, 1995. She was unable to complete additional testing

* Report contributed by Larry Beutler, PhD, Pacific Graduate School of Psychology, Palo Alto, CA.

during the intake because of "discomfort" and "apprehension" about making the drive home. She was, therefore, rescheduled with the request that she complete the rest of the paper-and-pencil materials over the course of the next two weeks.

During the total assessment sessions, she was evaluated using the following procedures:

Life History Questionnaire (12/12/95).
Clinical Interview (12/12/95).
Structured Clinical Interview for *DSM-IV* (SCID; 12/22/95).
State-Trait Anxiety Inventory (STAI; 12/13/95).
Millon Clinical Multiaxial Inventory-III (MCMI-III; 12/13/95).
Personal Attitude Inventory (Dowd Therapeutic Reactance Scale; 12/13/95).
Beck Depression Inventory (BDI; 12/13/95).
Beck Hopelessness Scale (BHS; 12/13/95).
Minnesota Multiphasic Personality Inventory-2 (MMPI-2; 12/21/95).
Sarason Social Support Questionnaire (SSQ; 1/2/95).

BEHAVIORAL OBSERVATIONS

Throughout the evaluation, A. G. was articulate, introspective, and cooperative. While acknowledging discomfort in talking about her sexual orientation and religious feelings, she was quite forthcoming when questioned directly. A valid MMPI-2 profile, along with her willingness to cooperate and introspect, suggests that the current evaluation presents a valid picture of her current level of functioning.

RELEVANT HISTORY

History of Presenting Problem. This 70-year-old, Jewish woman identifies three major symptom clusters that have been problematic for her: agoraphobia, panic attacks, and dissociation. She has a long history of panic attacks without agoraphobia, dating to age 12. The first panic episode occurred when she was babysitting for a family friend. She suddenly hyperventilated, began experiencing heart palpitations, and became afraid that she was going to die. She ran into the street yelling for help, but no one heard her or tried to assist her. The situation was resolved by exerting "self-control."

After her initial panic attacks began, they gradually increased to a frequency of about once per week throughout her teenage years. To protect herself from feared panic and what she perceived as possible death, she frequently slept with her parents and confined herself to known places and locations. At their worst, her panic attacks involved physical symptoms such as nausea, shortness of breath (hyperventilation), and dizziness, as well as cognitive symptoms such as fears of losing her mind, dying, of being overwhelmed, and unspecified danger. However, she learned to control these symptoms over time by avoiding such activities as going out, driving, and socializing with groups. These efforts have been successful in that A. G. reported that she had been asymptomatic for agoraphobia and panic for 31 years.

She currently reports that she has become apprehensive about travel and social activity, but the symptoms are confined to initial anticipatory anxiety, gastrointestinal distress, and headaches but with no heart palpitations, shortness of breath, or fainting. She prevents more extensive symptoms by avoiding travel and through a variety of distraction procedures. When she begins to experience the onset of panic, she calls someone or begins to read an interesting book. Her contacts with other people at these times do not include a disclosure of or discussions about the panic, but are reported to be simply methods to involve herself with others and to take her mind off her feelings.

She reported that the current symptoms are mild in intensity and include a general distaste of travel, an inability to get comfortable with being out alone, and a general heightened sense of vulnerability and apprehension until she is able to return home. She continues to avoid night travel and avoids being alone, if possible, to prevent the associated anxiety.

Since 1986, several dissociative episodes have occurred, which she believes were precipitated by her decision to openly acknowledge her homosexual orientation. The first instance followed a sexual encounter with her current partner during a vacation. After the sexual act, the client experienced an apparent fugue state. She became disoriented, was unable to recall personal information such as that her parents were deceased, and engaged in distraught communication with her lover about "Why am I here." The episodes have subsequently recurred several times: They come on suddenly and without warning and she subsequently has no memory for the events. They uniformly follow a lesbian sexual encounter, and if her partner remains with her during this period (sometimes up to several hours), the fears gradually subside. However, after these dissociative states, she reported having a sense of helplessness, hopelessness, confusion, headaches, and nausea that sometimes lasted for several days. She has been able to successfully avoid these episodes by not engaging in sexual activities for nearly six years.

History of Treatment. Ms. G. was first treated and hospitalized in 1965 because of agoraphobia. There have been no subsequent hospitalizations. However, she has entered into two treatment relationships in the years since. Her current medication is managed by a psychiatrist, who is treating her with Xanex, Tagamet, and Paxil. She reported that since being on the medication, she has been excessively sleepy and has a difficult time staying awake during the day. She also has experienced an increase in stomach difficulties and diarrhea. She was also treated for a short time in 1985 by an internist and psychiatrist. At that time, she was given tricyclic antidepressants. These drugs produced hallucinations and were discontinued shortly after initiation.

A review of this woman's symptom history also reveals that she has had substantial periods of time in which she has been asymptomatic for fugue states, panic attacks, and agoraphobia. She reported that between August 1994 and June 1995, she was the "best ever." She was able to travel alone, found life enjoyable, and experienced no episodes of discomfort or fear. More recently, she has gradually become more depressed and dysphoric as well as fearful, although there was no obvious precipitator for these feelings.

Family Background. A. G. was raised in a middle-class, Jewish family. She was the older of two children, having a brother who is one and a half years her junior. The

family always maintained at least a superficial religious identity and a facade of happiness. However, she reported that, behind this façade, there were significant underlying family conflicts. Religion has always been a source of conflict for her as has her sexual orientation. Moreover, she reported a long family history of mental illness and interpersonal conflict. While she described her parents as emotionally stable, both her parents' families have histories of psychiatric disorders. Her mother was the oldest of nine children; an uncle died in a halfway house with a diagnosis of schizophrenia, another was diagnosed as having bipolar disorder, and still another was diagnosed as having depression. On her father's side, at least one uncle is reported to have had a major depression that was treated with antidepressants.

In her own personal history, A. G. reported that she always felt confused about her sexual orientation. At age 17, she received a proposal of marriage. She declined but he persisted, and she went to live with an uncle to escape his advances. He pursued her and finally, against her "better judgment," he talked her into marriage. The newlyweds moved to Metroville to be with her family, but problems persisted and they separated after about a year. By that time, she had given birth to a daughter. She moved in with her parents, but long-standing conflicts with her mother became more frequent. When her husband contested and prevented the culmination of the divorce, A. G. moved out of the family home and went to work, leaving her daughter with her mother. A. G. blamed her parents for her failed marriage and refused further contact. She did not see or speak with her daughter for two years.

After a period of estrangement from parents and daughter, the patient was contacted by her attorney, who informed her that A. G.'s mother could no longer raise the baby. Her husband was also informed, and he demanded that A. G. reconcile with him to raise the child. A. G. agreed to do so if they would move to Betterville to make a fresh start. Shortly after moving, A. G.'s husband became disillusioned and returned to Metroville, leaving her to raise the child. It was very shortly thereafter that she acknowledged to herself that she was a lesbian. She subsequently engaged in a series of brief lesbian affairs, and adopted a "secret life" in which she prevented her parents and husband from an awareness of this emerging sexual orientation.

Still being unable to raise her daughter and work, the client gave up the child to a foster family for temporary care. After a few years, she initiated an effort to again assume care of the child. Concerned about raising the child in a lesbian relationship, she accepted the proposal of marriage from a man who knew about her lesbian lifestyle. He, nonetheless, agreed to adopt the child and allowed her to continue her lesbian affairs. Their marriage lasted 23 years and produced two sons. Although unsatisfied with her dual life, she waited until her younger son graduated from high school before she left the marriage and began to pursue lesbian relationships exclusively and openly. She met her current lover in 1986. This relationship continues to be close although they ceased sexual contact approximately six years ago to prevent dissociative episodes.

A. G. reported being close to her brother when she was young, but that relationship became disrupted during their teenage years. At that time, he had more problems with their parents than she did; and, because of these conflicts, A. G.'s parents left all their money to her rather than to her brother. However, she had come to feel alienated from him and subsequently refused to share her inheritance. In spite of having become quite wealthy, he became resentful of her refusal to share the inheritance and cut off the

relationship with her. In the past few years, she has tried unsuccessfully to reestablish this relationship. He has indicated his willingness to "forgive" her, but denies any desire for a continuing relationship.

Medical History. A. G.'s medical history is unremarkable. She currently takes Tagamet for stomach distress and Xanex for anxiety. Aside from some loss of hearing and psychophysiological symptoms, she acknowledges no significant medical problems.

INTERPRETATIONS AND IMPRESSIONS

While a formal assessment of intellectual level was not undertaken, both A. G.'s verbal conceptual skills and oral presentation suggest at least average and probably bright normal intellectual performance. Her ideation is dominated by preoccupation with ways to avoid uncomfortable feelings, along with concerns with physical symptoms. Collectively, this results in mild impairment to her cognitive efficiency. Her verbal processes are organized, circumstantial, and occasionally dominated by topics about which she has pressing concerns; but they reflect no disorganization, no memory impairment, and moderately well-developed associative and abstract reasoning processes. While she is oriented in all three spheres and manifests no significant mental impairment, she notes having always been concerned with the potential loss of mental functions.

A. G. denies dysphoria, depression, and anxiety. She complains of poor sleep, loss of energy, and lack of motivation. Formal assessment confirms the presence of vegetative signs (increased appetite, variable sleep, social withdrawal, loss of interest, reduced libido), consistent with the presence of mild to moderate depression without subjective dysphoria. Trait anxiety levels are within the normal range for her age, and subjective depression is only mild, with the dominant symptoms being psychophysiological. Her affect is appropriate, though somewhat variable. Affective responsivity is both dysthymic and blunted.

Ms. G.'s mood disturbance reflects a chronic condition, against which she has constructed a variety of rigid and brittle defenses. She is excessively sensitive to environmental signals of threat and, at the least suggestion of emotional arousal, engages in both direct and cognitive avoidance patterns. The result is that she prevents the intensification or even emergence of feelings that might overwhelm her. While protecting her somewhat from subjective sensations of anxiety and dysphoria, A. G.'s defenses are not sufficiently strong to prevent the emergence of a variety of secondary symptoms. Denial, phobic avoidance in the face of anticipatory cues, self-criticism, compartmentalization, and somatization are among her most frequently used defenses. As threat intensifies, her fragile denial deteriorates, and both somatization and direct avoidance predominate. Thus, acute stress evokes a variety of stress-related somatic symptoms and phobic behaviors that provide expression for her denial of anxiety and depression.

A. G. experiences ambivalent personality organization, with moderate disturbances to her functional adaptation. Her dominant conflicts involve strong needs for dependency, counterbalanced by equally strong strivings for self-definition. Her coping style involves both passive and active efforts to reconcile these strong drives. Thus, while she seeks approval and confirmation from others, even to the point of excessive subservience in which she gives up personal strivings, this is frequently a futile effort to ensure the

presence of other people in her life. Indeed, these efforts are usually designed to compensate for a host of covert rebellious and angry impulses and by overt efforts to be autonomous and self-guided. Thus, efforts to achieve self-fulfillment and autonomy are followed by guilt, self-doubt, and shame in which fear and withdrawal dominate. These latter symptoms, however, may be so demanding of attention that they are the functional equivalent of interpersonal anger, hurt, and resentment. Thus, her pattern of phobic anxiety and dissociation has led to sexual withdrawal and physical dependency. This may allow an indirect expression of anger, yet also be a compromise between asocial impulses and needs for approval. Unfortunately, this compromise also includes low self-regard and restricted mobility. Another consequence of this pattern is the current low level of available others to provide support. In spite of this, A. G.'s satisfaction with the level of interpersonal support available from her significant other is good and suggests the availability of this individual as a support in any treatment program.

DIAGNOSTIC IMPRESSION

AXIS I	(300.22)	Agoraphobia without recent panic disorder (by history)
	(296.3)	Rule out Major Depressive Disorder, recurrent
		The diagnosis of Anxiety Disorder is based on history rather than current symptomatology. A major differential question has to do with the relevant salience of Major Depression versus Anxiety Disorder
AXIS II		None
AXIS III		Rule out gastrointestinal disorder
AXIS IV		Problems related to social environment-social isolation, restriction of friendships, lifecycle transitions
AXIS V		Current GAF: 62 Highest GAF past year: 80

SUMMARY AND RECOMMENDATIONS

Overall, the foregoing confirms the presence of a complex and long-standing problem founded more in dynamic and early developing interpersonal expectations and conflicts than in symptom-contingent events. However, this is not meant to discount the importance and debilitating nature of the patient's symptom picture, nor the need to give it direct attention in treatment. The dynamic nature of the associated conflicts, and their role in maintaining systemic dysfunction in her relationships with significant others, suggests the need to combine a symptom-focused treatment with efforts to resolve fundamental conflicts. These core conflicts seem to be largely founded in postpubescent strivings to resolve needs for autonomy and dependency. The initial focus of treatment should be on reducing territorial apprehension, with a concomitant increase in social involvement and independent functioning. After initial symptomatic improvement, further interventions should focus on A. G.'s pattern of rebelliousness, which seems to be intertwined with self-incrimination, guilt, and withdrawal. In particular,

this might emphasize confirming needs for both autonomy and acceptance, along with greater insight into this pattern.

A. G.'s level of functional impairment is moderate. Numerous areas of functioning are affected, and this, coupled with the chronicity of the condition, suggests the need for long-term treatment. However, the level of defense and personal control is sufficient, and the level of subjective despair and hopelessness is within a range that suggests that outpatient care is appropriate. There is no evidence of direct risk to self or others. Anxiolytic or antidepressant medications are contraindicated because of her relative degree of control over her symptoms, combined with the high potential for somatic side effects. Individual treatment may allow a more selective and intensive focus on problematic behaviors. Individual therapy would also be likely to prevent the operation of direct avoidance of discomfort when compared with group treatment.

A. G.'s level of distress is well contained since it falls in the average or even below-average range compared with other patients who seek treatment. While her distress increases significantly when exposed to immediate threat, she quickly compensates and is so well versed at avoidance that she may experience difficulty sustaining sufficient motivation for therapy. Thus, interventions that confront or expose her to feared and avoided circumstances may be helpful to desensitize her to anxiety as well as to increase her level of motivation to continue treatment.

A. G.'s coping style vacillates between being primarily impulsive and externalizing, to being self-critical and internalizing. This pattern of cyclic coping suggests the need to address her problems at both a behavioral and an insight level. When her impulses and direct avoidance dominate, behavioral strategies should be emphasized. During phases in which she is more introspective and self-blaming, insight-oriented interventions are likely to be more effective. Given the unsustaining nature of her subjective distress, abreactive and sensate-focused, cathartic interventions may prove to be especially helpful during these more introspective phases.

Finally, A. G. manifests a pattern of superficial compliance and more covert resistance to the directives of help givers. Thus, special attention should be given to developing a trusting relationship. Even if this is achieved, however, she would still be expected to undermine direct suggestions and specific assignments. The most effective approaches, then, would be collaborative interventions emphasizing clear behavioral change, contingency contracting, or paradoxical interventions such as symptom prescription and "no-change" directives. Particular attention may be given to predicting the exacerbation of physical and phobic symptoms following intense sessions because these sessions may mobilize her resistant impulses in an asymptomatic direction.

Collectively, the symptomatic aspects of the patient's fears and phobias may be susceptible to a combination of structured exposure procedures, cognitive restructuring, and interoceptive awareness (Craske & Barlow, 1993). These procedures circumvent patient resistance by virtue of their reliance on self-monitoring, as well as being both symptom and behaviorally focused. The more thematic and dynamic aspects of A. G.'s problem may be addressed by initiating work that specifically mobilizes her anxiety in motivational directions. Confrontation with feared material, along with the use of procedures such as two-chair work and imaginal reliving of unsettling relationships, may be helpful. Imaginal confrontation might be initiated with images and memories of disapproving parents, children, and other significant others, the goals of which may be to

help her tolerate discomfort and disapproval. The procedures outlined by Daldrup, Beutler, Engle, and Greenberg (1988, *Focused Expressive Psychotherapy*) for working with the overcontrolled patient may also be particularly helpful.

RECOMMENDED READING

Armengol, C. G., Kaplan, E., & Moes E. J. (Eds.). (2001). *The consumer-oriented neuropsychological report.* Lutz, FL: Psychological Assessment Resources.

Tallent, N. (1993). *Psychological report writing* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Zuckerman, E. L. (2000). *The clinician's thesaurus: A guidebook for wording psychological reports* (5th ed.). Pittsburgh, PA: Three Wishes Press.

Appendix A

TEST PUBLISHERS/DISTRIBUTORS

American Guidance Service, Inc.
Publisher's Building
Circle Pines, MN 55014-1796
1-800-328-2560
List includes: Kaufman Assessment Battery for Children, Kaufman Adolescent and Adult Intelligence Test, Kaufman Brief Intelligence Test

American Orthopsychiatric Association, Inc.
19 West 44th Street
New York, NY 10036
(212) 354-5770
List includes: Bender Visual-Motor Gestalt Test

Boston Neuropsychological Foundation
P.O. Box 476
Lexington, MA 02173
List includes: Stroop Color-Word Test (Comali/Kaplan version)

Center for Behavioral Health Care Technologies, Inc.
3600 S. Harbor Blvd. #86
Oxnard, CA 93035
(805) 677-4501
www.systematictreatmentselection.com
email: info@cbhti.com
List includes: Systematic Treatment Selection software

Consulting Psychologists Press
3803 East Bayshore Road
P.O. Box 10096
Palo Alto, CA 94303
1-800-624-1765
(650) 969-8901
www.cpp-db.com
www.SkillsOne.com
List includes: California Psychological Inventory, Fundamental Interpersonal Relations Orientation-B, Myers Briggs Type Indicator, Strong Interest Inventory