FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Evaluating User Engagement in Internet-Based Behavioral Therapies and Its Impact on Treatment Outcomes

Rodrigo Manuel Graça Figueiredo

Master in Informatics and Computing Engineering

Supervisor: Carla Teixeira Lopes

July 29, 2026

Evaluating User Engagement in Internet-Based Behavioral Therapies and Its Impact on Treatment Outcomes

Rodrigo Manuel Graça Figueiredo

Master in Informatics and Computing Engineering

Approved in oral examination by the committee:

President: Prof. Sérgio Nunes

Examiner: Prof. Dagmara Paiva

July 29, 2026

Resumo

As perturbações da saúde mental representam um importante desafio global de saúde pública, afetando centenas de milhões de pessoas em todo o mundo. As intervenções comportamentais baseadas na Internet, disponibilizadas através de plataformas digitais, têm emergido como uma abordagem escalável para expandir o acesso ao apoio psicológico. Embora o envolvimento dos utilizadores seja amplamente reconhecido como um fator determinante da eficácia destas intervenções, os métodos sistemáticos para medir o envolvimento diretamente a partir de dados de interação continuam pouco desenvolvidos na área da saúde mental digital.

Esta dissertação investiga de que forma o envolvimento dos utilizadores numa plataforma de terapia comportamental baseada na Internet pode ser quantificado através de dados de registo de interações e como os padrões de envolvimento se relacionam com os resultados terapêuticos. O estudo analisa dados recolhidos durante uma intervenção de terapia cognitivo-comportamental com duração de quatro semanas, realizada através da plataforma Moodbuster, composta por uma aplicação móvel destinada à avaliação ecológica momentânea (EMA), através da qual os participantes recebiam notificações diárias que os convidavam a avaliar o seu humor, ansiedade e qualidade do sono, e por um portal web com módulos terapêuticos.

Foi realizada uma revisão sistemática da literatura para identificar abordagens existentes à avaliação do envolvimento baseada em registos de interação em diferentes plataformas digitais, com especial foco em contextos relacionados com a saúde. Com base nesta revisão, foi desenvolvido um quadro multidimensional de métricas de envolvimento que abrange seis dimensões: tempo de utilização, volume e frequência de interação, acesso e navegação em conteúdos, conclusão de tarefas e desempenho, comportamento de utilização ao longo do tempo e tecnologia de acesso.

Foram calculadas métricas de envolvimento a partir de registos de interação que incluíam notificações, avaliações de humor, registos em texto livre, comportamento de navegação, progressão nos módulos, atividade de início de sessão e comunicação com terapeutas. A análise revelou padrões distintos de utilização entre as duas plataformas: o envolvimento na aplicação móvel consistiu principalmente em sessões curtas e frequentes centradas no registo do humor e influenciadas por notificações, enquanto o envolvimento no portal web refletiu uma progressão estruturada através dos módulos terapêuticos, com diminuição das taxas de conclusão nas fases mais avançadas.

As associações entre as métricas de envolvimento e as alterações na motivação, avaliada através da escala SMFL (Short Motivation Feedback List) e na gravidade dos sintomas depressivos, avaliada através da subescala HADS-D (Depression subscale of the Hospital Anxiety and Depression Scale), foram avaliadas através da correlação de Spearman. As relações observadas foram, de modo geral, modestas. A utilização sustentada da aplicação móvel apresentou associações positivas fracas com a redução dos sintomas, enquanto uma participação consistente e persistente no portal web mostrou associações mais fortes com melhorias na motivação do que períodos isolados de utilização intensiva. Modelos de regressão logística com regularização L2

apresentaram desempenho próximo da classificação aleatória, sugerindo que os dados de interação, por si só, são insuficientes para prever de forma fiável os resultados terapêuticos.

Os resultados sustentam a conceptualização do envolvimento como um conjunto multidimensional e dependente do contexto nas intervenções digitais em saúde mental: o mesmo comportamento assumia significados diferentes dependendo da plataforma, uma vez que as métricas informativas para a aplicação móvel não eram necessariamente relevantes para o portal web, e vice-versa. Dentro deste quadro dependente do contexto, um padrão manteve-se consistente em ambas as plataformas: a consistência e a regularidade da participação pareceram ser mais relevantes para os resultados terapêuticos do que o volume global de atividade. Este trabalho demonstra a viabilidade de extrair conhecimentos comportamentais relevantes a partir de registos de interação e estabelece uma base para futuros modelos de análise do envolvimento em intervenções digitais de saúde mental.

Palavras-chave: saúde mental • terapia comportamental • envolvimento • análise de logs • digital

Abstract

Mental health disorders represent a major global public health challenge, affecting hundreds of millions of people worldwide. Internet-based behavioral interventions delivered through digital platforms have emerged as a scalable approach to expanding access to psychological support. Although user engagement is widely recognised as a key determinant of intervention effectiveness, systematic methods for measuring engagement directly from interaction data remain underdeveloped in digital mental health research.

This dissertation investigates how user engagement in an internet-based behavioral therapy platform can be quantified using interaction log data and how engagement patterns relate to therapeutic outcomes. The study analyses data collected during a four-week cognitive behavioral therapy intervention delivered through the Moodbuster platform, consisting of a mobile application supporting ecological momentary assessment (EMA) through which participants received daily notifications prompting them to rate their mood, anxiety, and sleep quality, and a web portal containing structured therapeutic modules.

A systematic literature review was conducted to identify existing approaches to log-based engagement assessment across digital platforms, with emphasis on health-related contexts. Based on this review, a multidimensional engagement framework was developed encompassing six dimensions: time spent, interaction volume and frequency, content access and navigation, task completion and performance, longitudinal usage behavior, and technology of access.

Engagement metrics were calculated from interaction logs capturing notifications, mood ratings, free-text entries, navigation behavior, module progression, login activity, and therapist communication. Analysis revealed distinct usage patterns across platforms: mobile application engagement consisted primarily of brief, frequent sessions centered on mood reporting and influenced by notifications, whereas web portal engagement reflected structured progression through therapeutic modules with declining completion in later stages.

Associations between engagement metrics and changes in motivation, assessed with the Short Motivation Feedback List (SMFL), and depressive symptom severity, assessed with the Depression subscale of the Hospital Anxiety and Depression Scale (HADS-D) were assessed using Spearman's rank correlation. Relationships were generally modest. Sustained mobile application use showed weak positive associations with symptom reduction, while consistent and persistent participation in the web portal was more strongly associated with motivational improvement than isolated periods of intensive use. Logistic regression models with L2 regularisation achieved near-random predictive performance, suggesting that interaction logs alone are insufficient for reliable prediction of therapeutic outcomes.

These findings support conceptualising engagement as a multidimensional and context-dependent construct in digital mental health interventions: the same behavior carried different meaning depending on the platform, since metrics that were informative for the mobile application were not necessarily meaningful for the web portal, and vice versa. Within this context-dependent picture, one pattern held across both platforms: consistency and regularity of participation appeared more relevant to therapeutic outcomes than overall activity volume. This work demonstrates the feasibility of deriving meaningful behavioral insights from interaction logs and provides a foundation for future engagement analytics frameworks in digital mental health interventions.

Keywords: mental health • behavioral therapy • engagement • log analysis • digital

The United Nations Sustainable Development Goals

The United Nations Sustainable Development Goalss (SDGs) provide a global framework for achieving a better and more sustainable future for all. It includes 17 goals (https://sdgs.un.org/goals) to address the world's most pressing challenges, including poverty, inequality, climate change, environmental degradation, peace, and justice.

The specific Sustainable Development Goal addressed in this work has the following name:

SDG 3 Ensure healthy lives and promote well-being for all at all ages.

SDG	Target	Contribution	Performance Indicators and Metrics
3	3.4	Understanding how user engagement relates to the efficacy of digital therapy, developing several metrics that may capture this engagement, supporting a better understanding of how patients interact with digital therapies and what behaviors are related to better therapy outcomes.	Higher correlation between user engagement and reduction in depression.
	3.8	Analyzing a digital therapy platform and how patients interact with it, with insights into what can be improved, supporting the development of improved new digital therapy platforms.	Increase in the number of digital therapy platforms and corresponding usage.

Acknowledgements

I would like to express my sincere gratitude to everyone who supported me throughout the completion of this dissertation.

First and foremost, I would like to thank my family for their encouragement, patience, and support. Their confidence in me has been a constant source of motivation throughout this journey.

I am especially grateful to my supervisor for her guidance, encouragement, and continuous support throughout the entire process. Her dedication and willingness to stand by me, even during periods when the circumstances were far from ideal, were invaluable and greatly contributed to the completion of this work.

I would also like to thank INESC TEC for providing me with a research grant (11554/BI-L-EM_B3/2025) during the development of this dissertation, which gave me the opportunity to contribute to a European research project. I am equally grateful for the access to the dataset that served as the foundation for this work and made this research possible. [1]

Finally, I would like to thank everyone who, in one way or another, contributed to this work and made this journey both possible and rewarding.

Rodrigo Figueiredo

[1]This work was co-financed by Component 5 - Capitalization and Business Inovation, integrated in the Resilience Dimension of the Recovery and Resilience Plan within the scope of the Recovery and Resilience Mechanism (MRR) of the European Union (EU), framed in the Next Generation EU, for the period 2021-2026, within project HfPT, with reference 41. The data used in this work was obtained in the context of a long-standing collaboration between INESC TEC and Vrije Universiteit Amsterdam (VUA), currently carried out within the RECONNECTED project, funded by the European Union under the Horizon Europe programme (Grant Agreement No. 101081020)

Declaration of honour

I declare that this dissertation is my own work and has not previously been submitted or assessed at this or any other institution. References to other authors (statements, ideas, thoughts), as well as the use of published works, strictly comply with the rules of attribution and are duly identified in the text and in the bibliographic references, in accordance with the applicable referencing standards. I am aware that the practice of plagiarism or self-plagiarism constitutes an academic offence, as established in the U.Porto Code of Ethics.

I also declare that I have used Artificial Intelligence tools to complete part(s) of this dissertation, and that this use and its results are duly noted and have been carefully checked for errors, inaccuracies, unfounded attributions or other forms of bias in content and arguments, as well as determining authorship.

Rodrigo Manuel Graça Figueiredo

Rodrigo Figueiredo

Contents

List of Figures

List of Tables

List of Acronyms

AUC-ROC Area Under the Receiver Operating Characteristic Curve
CBT Cognitive Behavioral Therapy
CSQ-I Client Satisfaction Questionnaire for Internet-based Interventions
CSV Comma-Separated Values
DMHI Digital Mental Health Intervention
EMA Ecological Momentary Assessment
HADS-D Hospital Anxiety and Depression Scale – Depression subscale
IBI Internet-Based Intervention
IHME Institute for Health Metrics and Evaluation
IQR Interquartile Range
mHealth Mobile Health
PHQ-9 Patient Health Questionnaire – 9 items
PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SDG Sustainable Development Goals
SMFL Short Motivation Feedback List
SUS System Usability Scale
WHO World Health Organization

Chapter 1

Introduction

1.1 Context and Motivation

Despite the critical importance of mental health in overall well-being, disorders such as anxiety, depression, and stress-related disorders are increasingly common, impacting hundreds of millions of individuals [33]. According to the Institute for Health Metrics and Evaluation (IHME) [33], the COVID-19 pandemic significantly exacerbated this issue. In 2020, 18% more people experienced depressive disorders and 15% more experienced anxiety disorders compared to 2019 [33]. In 2023, 15% of the world's population experienced mental disorders [33].

These mental health issues can influence various aspects of life, including relationships with family, friends, and the community, as well as academic and occupational performance. According to the World Health Organization (WHO) [72], mental disorders are among the top 10 leading causes of health loss worldwide, with anxiety and depressive disorders ranked as the most burdensome across all age groups and locations. Between 1990 and 2023, mental disorders jumped up in the ranking of top causes of health loss worldwide, from 12th to 5th place. Globally, mental disorders account for 1 in 6 years lived with disability, and individuals with severe mental health problems typically have a lifespan 10 to 20 years shorter than the average population [72]. In addition to personal and societal effects, the economic impact is substantial, with the decrease in productivity exceeding the treatment expenses [38, 46, 72].

Health systems are still lacking resources to a great extent, and there is a large treatment gap all over the world [72]. Anxiety disorders provide a striking example, with an estimated 71% of the global burden being avoidable if all affected individuals had access to optimal treatment. To help bridge this gap, in recent years, there has been significant growth in the development, evaluation, and availability of Internet-based interventions aimed at treating behavioral and mental health disorders [3, 7]. Furthermore, there has also been a rapid increase in health-related mobile and desktop applications to aid in treating chronic health conditions [13]. These digital mental health interventions can expand access to mental health resources and support while removing some of the current obstacles to traditional care, such as reducing stigma through private and anonymous use, offering affordable alternatives to in-person therapy, increasing the availability of evidence-based tools regardless of location, and providing support to individuals who may not have a formal diagnosis but still experience mental health challenges. [8].

1.2 Problem Statement

Research has consistently shown that higher levels of user engagement are associated with the success of a system [32]. One notable example in the field of digital mental health is the RECONNECTED project [53], a European initiative aiming to improve citizens' mental health and well-being. Central to this initiative is the Moodbuster platform, designed to provide evidence-based psychological treatments for mental health conditions.

However, despite the potential of leveraging user engagement to improve the success of health-related technologies and interventions, a substantial gap persists in the systematic use of engagement measurement and predictive methods. While extensive research has been conducted in domains such as education, few studies explore user engagement analysis and its impact on the success of therapies.

Current methods for measuring user engagement, such as questionnaires, are valuable but indirect and subjective, offering only a limited view of the complexity of user behavior on digital platforms. This limitation restricts the potential for optimizing digital health interventions to maximize effectiveness.

To address this gap, this work investigates whether user engagement in digital mental health interventions can be measured directly from interaction data and how such measures relate to indirect measures such as questionnaires, as well as therapy outcomes. In this study, we developed metrics based on detailed user activity logs and evaluated their potential to capture meaningful engagement patterns. Additionally, predictive models were developed to identify specific user behaviors indicative of positive therapy outcomes, ultimately providing insights to enhance user engagement and therapy effectiveness. These methods were applied and validated using data from the Moodbuster platform.

1.3 Background

This section starts by providing an overview of existing Digital Mental Health Interventions (DMHIs), highlighting their main formats, features, objectives, and associated challenges. Then, it presents the Moodbuster platform in more detail, including its structure and functionalities, as it will be used as a case study.

1.3.1 Digital Mental Health Interventions (DMHIs)

Digital mental health interventions (DMHIs) are psychological interventions that have been adapted or translated into digital formats, designed to enhance access to psychological support and improve the efficacy, effectiveness, and efficiency of care. [6]. Their main aims are to widen access to care and close treatment gaps tied to cost, geography, or provider shortage [6, 41]. They also support prevention and early intervention, reduce stigma, and reach people below diagnostic thresholds or reluctant to seek formal help [6, 39, 41, 48]. Consequently, the telemedicine and health-app market is expanding rapidly, a trend accelerated by COVID-19 [41]. Smartphone apps now span the entire continuum of care, from prevention and diagnosis to adjunctive treatment and relapse prevention, offering low-intensity, cost-effective, and highly accessible support [6]. Economic evaluations generally show a favorable cost-effectiveness profile [41].

The most well-established formats include computerized cognitive behavioral therapy (CBT) [39], mobile mental health apps (mHealth) [6, 8, 39, 41], and guided self-help platforms [6]. These are complemented by more innovative interventions such as therapeutic video games [8, 39, 48], wearable sensor-based monitoring tools [8, 41, 48], virtual reality (VR)-based exposure therapies [6, 41, 48], and AI-driven conversational agents or chatbots [6, 41].

DMHIs often offer functional capabilities such as real-time symptom monitoring [8, 39, 41], behavioral feedback [6], educational information [8, 39], and adaptive, personalized support [6]. For instance, wearables and apps can monitor sleep and activity patterns [41]; VR systems provide immersive therapeutic experiences [41]; and chatbots can engage in natural language conversations and simulate human-like behavior during goal-directed interactions with people [41], with some studies already showing that their effectiveness is on par with traditional face-to-face therapy [6]. Other commonly integrated features include counseling, mindfulness exercises, peer support, and text messaging functions such as reminders [8].

A prominent format among DMHIs is Internet-Based Interventions (IBIs), which are one of the most extensively researched. IBIs are structured e-health treatments delivered through a computer or mobile device. They are organised into sequential modules or lessons, each targeting a clearly defined psychological goal [6] - an approach very similar to how the Moodbuster platform is set up. Evidence indicates that IBIs can match the effectiveness of conventional in-person therapy across numerous mental health conditions, and they are also effective in promoting healthy behaviors and preventing psychological problems [6]. These interventions often combine with or serve as an alternative to in-person care, particularly useful in prevention, diagnosis, treatment, and relapse management [6].

1.3.2 Moodbuster Platform

Moodbuster is a digital platform developed as part of the RECONNECTED project [53], specifically designed to support individuals experiencing depression through structured, evidence-based interventions. It offers a range of online cognitive behavioral therapy modules accessible via web and mobile applications.

Moodbuster consists of a web portal and a mobile app, each serving complementary functions. The portal provides access to therapy modules and therapist interaction, while the app supports real-time monitoring and daily engagement. Together, they offer an integrated digital therapy experience.

1.3.2.1 Web Portal

The web portal is the primary interface for both patients and therapists. For patients, it provides access to various therapy modules that combine educational lessons, interactive exercises, and take-home assignments designed to tackle multiple dimensions of depression. Through the portal, therapists can monitor patients' progress, review their completed modules, and communicate with them through a built-in messaging system. Figure 1.1, taken from a study done in the Netherlands, illustrates the portal's module overview page, where patients can check which modules they can access and launch anytime.

Figure 1.1: Moodbuster web portal - Patient's module overview. [3]

1.3.2.2 Mobile Application

The mobile application complements the web portal with real-time monitoring through Ecological Momentary Assessment (EMA). Patients rate their mood, anxiety levels, and sleep quality each day on a 1-to-7 scale, and they can also write freely about what they have eaten, how well they slept, and whatever thoughts are on their mind at that moment. Therapists can review the numerical check-ins and these text entries to fine-tune treatment plans. Figure 1.2 shows the rating screen where users can submit daily scores and access the history tab, represented in Figure 1.3, which displays a chart with the ratings given for a selected day, week, or year.

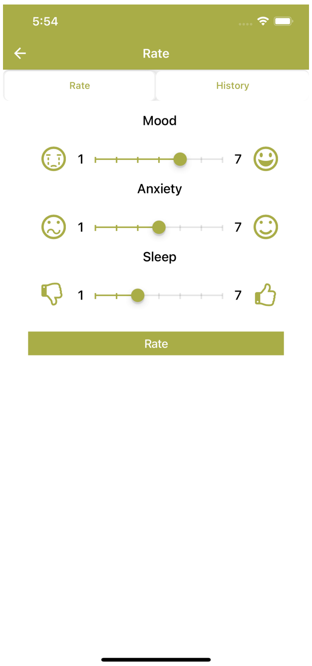


Figure 1.2: Moodbuster mobile app—daily rating interface. [3]

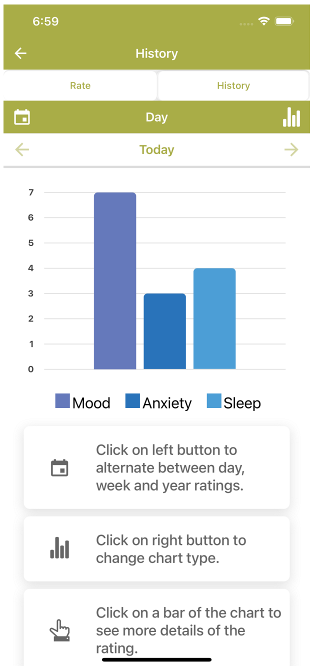


Figure 1.3: Moodbuster mobile app—rating history view.

1.4 Research Questions

To explore the relationship between user engagement and therapy success, this work analyzes direct interaction data from the Moodbuster platform to complement existing indirect engagement measures. The following research questions guide this study:

- **RQ1** : How do users interact with a digital mental health platform, as reflected in access log data?
- **RQ2** : What metrics can be defined to quantify user engagement?
- **RQ3** : How do direct engagement metrics from interaction data correlate with indirect measures of user engagement?
- **RQ4** : What is the relationship between user engagement and therapy outcomes?
- **RQ5** : What design improvements can be made to digital mental health platforms to enhance user engagement and therapy effectiveness, based on insights from user interaction data?

Chapter 2

Assessing user engagement and behavior through log file analysis - A systematic literature review

This chapter conducts a systematic literature review to map the state of the art in log-based assessment of user engagement and behavior, focusing primarily on methods and metrics relevant to digital health platforms, while also drawing on adaptable and generalizable insights from other domains.

2.1 Rationale

Most digital applications capture event logs, which are usually time-stamped records of user actions and system events. When analyzed, these logs can be crucial in understanding user interactions and engagement within those digital platforms. By examining the patterns within the logs, it is possible to gain insights into user behavior that are critical for enhancing platform effectiveness. In health-focused platforms, such insights are potentially valuable for improving patient outcomes.

However, research on engagement metrics and methods tailored to health platforms remains limited. This literature review adopts a broader approach to address this gap, examining methodologies and metrics from various domains. While some techniques may directly apply to health platforms, others may require adaptation to align with the specific goals and constraints of the health sector. This inclusive approach ensures a comprehensive understanding of existing strategies, providing a foundation for identifying and adapting the most relevant methods for health-focused platforms.

2.2 Methodology

A systematic review was conducted to identify and evaluate the current state of research on this topic. This review follows the PRISMA guidelines, systematically selecting and analyzing studies

Table 2.1: Query and number of results for each database.

Database	**Query**	**Results**
Google Scholar	"log file analysis" AND (interaction OR engagement) *(Filtered: Results from 2019 onward)*	994
Scopus	"log file analysis" AND (interaction OR engagement)	686
ACM Digital Library	[All: "log file analysis"] AND [[All: interaction] OR [All: engagement]]	216
Web of Science	"log file analysis" AND (interaction OR engagement)	38

investigating log data analysis about user engagement, behavior, and platform success.

This systematic review aims to answer the following research question:

- **RQ1** : What metrics quantify user engagement and behavior in digital platforms?

2.2.1 Information Sources

To ensure a comprehensive collection of literature, four major academic databases were chosen, each providing a vast repository of research papers: Google Scholar, Scopus, Web of Science, and ACM Digital Library.

2.2.2 Search Strategy

Several initial search queries were tested across all databases to determine the optimal balance between specificity and breadth. Queries that were too broad yielded excessive results, many of which fell outside the desired scope, while highly specific queries risked omitting potentially valuable studies. The queries were validated by analyzing the first pages of the results and checking if most of the papers could be relevant. After iterative refinement, a single query was selected as it consistently returned an appropriate and manageable number of papers across all sources. All databases were last searched on 26/10/2024 and given the large volume of results retrieved from Google Scholar, the search was restricted to publications from 2019 onwards. The query and number of results for each database are presented in Table 2.1.

2.2.3 Eligibility Criteria

To ensure a relevant selection of papers, specific inclusion and exclusion criteria were defined. For a paper to be selected, it must meet at least one of the inclusion criteria and only needs to meet one of the exclusion criteria to be excluded.

Inclusion Criteria

- **Techniques for Analyzing Log Data**: Paper explicitly discusses techniques, frameworks, or tools for analyzing log files to extract insights into user behaviors or engagement patterns.
- **Metrics for Engagement and Behavior**: Paper defines or implements specific metrics to quantify user engagement or behavior based on log data.
- **Linking Engagement with Outcomes**: Paper investigates the relationship between user engagement and platform success, effectiveness, or other measurable outcomes.

Exclusion Criteria

- **Not Focused on Active User Engagement:** Paper focuses on platforms primarily designed for providing data (e.g., health record portals) or for monitoring and evaluating technical system performance (e.g., anomaly detection). These platforms are not associated with interactive engagement or producing behavioral outcomes.
- **Non-Adaptive Contexts**: Paper focuses on domains or platforms where user engagement behaviors differ fundamentally from those in the health context, making the metrics non-transferable. For example, user interactions on transactional platforms like e-commerce differ from those in health-related platforms, such as therapy applications, as they serve distinct purposes and involve different engagement patterns.
- **Data Acquisition or Logging Only**: Paper focuses solely on collecting, storing, or managing log data without utilizing it to analyze user engagement or derive behavioral insights.
- **Survey or Questionnaire-Based Studies**: Paper relies primarily on self-reported data (e.g., questionnaires or surveys) rather than utilizing log data to measure engagement.
- **Usability-Only Studies:** Paper is focused solely on usability or user satisfaction without addressing engagement metrics or patterns.
- **Language:** Paper is not written in English.

2.2.4 Selection Process

A step-by-step selection process was followed to find the most relevant studies. First, papers were screened based on their titles and abstracts to exclude those that didn't fit the review's focus. Then, a deeper content assessment was done to ensure the final selection met all criteria. This approach helped narrow the most valuable studies while maintaining a structured and reliable selection process.

Figure 2.1 shows the PRISMA-style flow diagram that tracks each stage of the selection procedure. A total of 1,934 records were initially retrieved across four databases. After title and abstract screening and full-text eligibility checks, 54 articles remained for inclusion in the review.

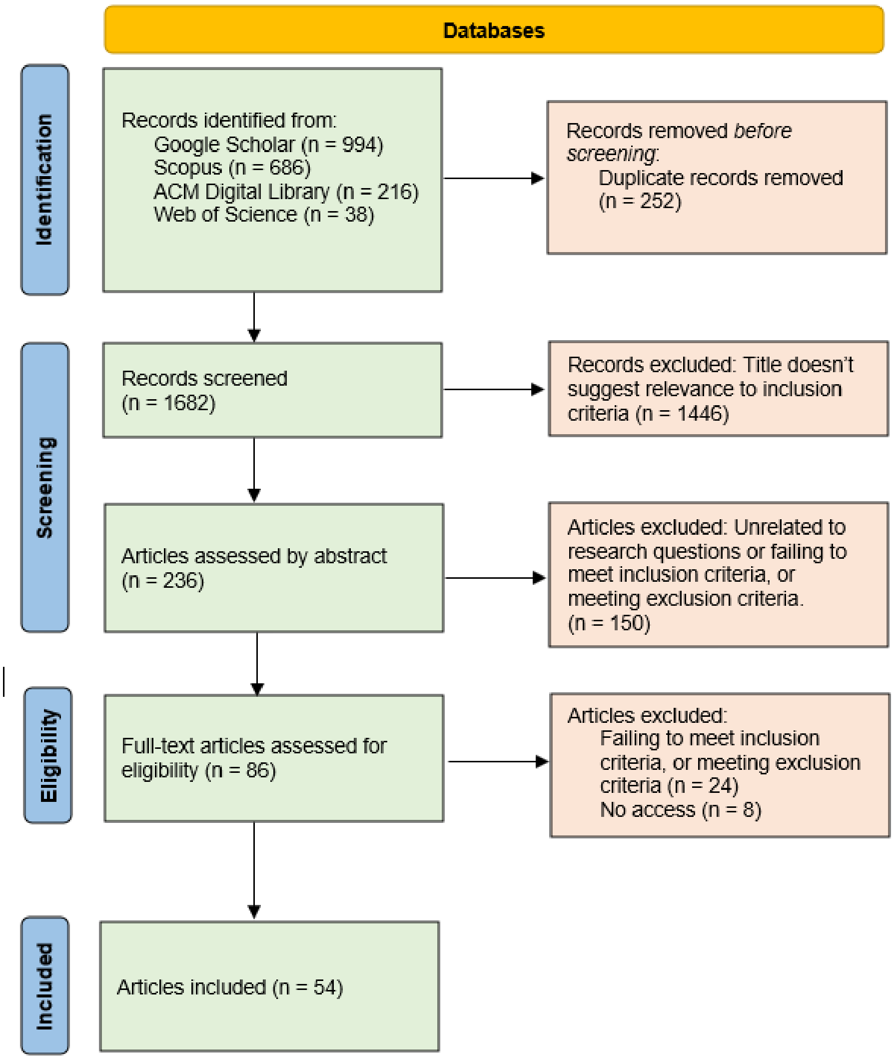


Figure 2.1: Flow Diagram providing an overview of the selection process.

After the database querying, all results were screened by the title and abstract. In the title screening phase, papers were evaluated to exclude studies that fell outside the scope of this review. Papers were included in the next phase if their titles suggested relevance to any of the inclusion criteria. Also, if it was evident from the title that a paper fulfilled one or more of the exclusion criteria, it was excluded. In cases where the title's relevance was uncertain or ambiguous, the paper was retained for further evaluation in the following phase, ensuring that potentially valuable studies were not prematurely excluded. From the 1,682 non-duplicate titles, 236 were selected for the next phase.

During the abstract screening phase, further refinement was applied to ensure that the selected papers met the inclusion criteria and didn't meet any exclusion criteria. After this phase, 86 papers were selected for further review. In both phases, results in all formats were accepted if they were accessible.

All papers that progressed beyond the abstract screening were reviewed for eligibility. This process evaluated their compliance with the inclusion and exclusion criteria while extracting key information from each paper. The goal was to comprehensively analyze the methods employed and identify existing gaps. In this step there were also eight articles that could not be included in the review because their full texts were not accessible through free or institutional access. Following this assessment, 54 papers were included.

2.3 Overview of Included Papers

The papers included after the selection process [1, 2, 4, 5, 9–12, 14–31, 34–37, 40, 42–45, 47, 49–52, 54–56, 58–70, 73, 74] focused on various approaches to analyzing log file data, extracting user behavior and engagement patterns, quantifying engagement through specific metrics, and predicting user engagement, performance, and the success of platforms or interventions.

2.3.1 Application Domain

Understanding the domain of the studies included in this review is critical to contextualize their findings and assess their adaptability to health platforms. The distribution of domains among the included papers is depicted in Figure 2.2.

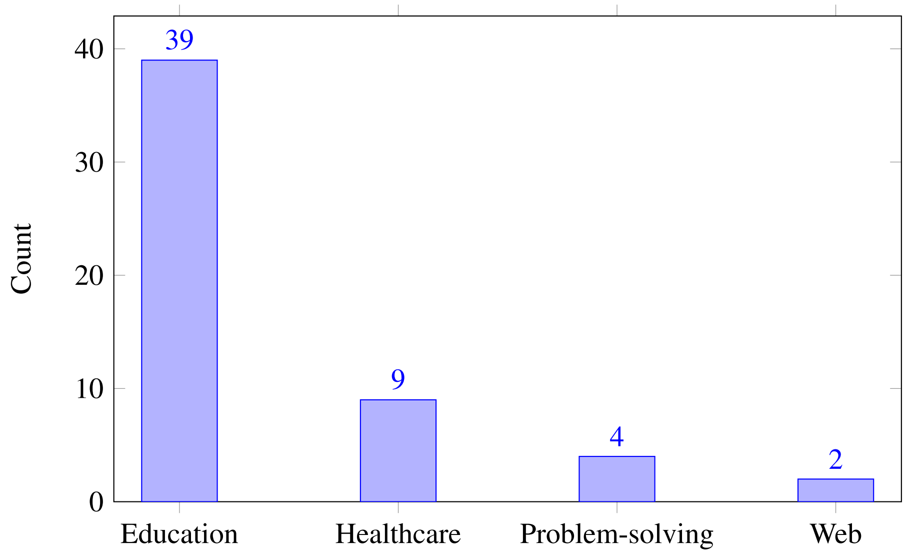


Figure 2.2: Distribution of included papers by domain.

As illustrated, most studies are from the education domain, with only nine papers from the health domain. While the number of studies in the health domain is significantly lower than in the education domain, many of the methodologies and metrics developed for educational platforms can be transferable or adaptable to healthcare systems. Both domains share key characteristics, and the measure of engagement may indicate effectiveness in both. Nevertheless, the disparity between the two domains regarding the number of papers found highlights the necessity for more targeted research for health-related platforms.

2.3.2 Metrics for Engagement Measurement

To answer Research Question 1, "What metrics quantify user engagement and behavior in digital platforms?", we extracted every engagement-related metric reported in the reviewed studies. Because the review spans health-focused and non-health-focused literature, the metrics are presented in two different tables: Table 2.2 lists the metrics employed in health-related studies, while Table 2.3 lists those used in non-health contexts. For clarity, metrics were grouped into six dimensions:

- **Time-on-Platform Measures**: Captures how long users interact with the platform or its features, as well as temporal patterns such as punctuality or inactivity.
- **Interaction Volume and Frequency**: Reflects how often and how intensively users interact with the platform, from logins and raw action counts to specific frequencies of engagement across tasks or time periods.
- **Content Access and Navigation**: Describes the breadth, depth, and revisitation of content, including page views, reading behaviors, and diversity of accessed material.
- **Task and Performance Outcomes**: Measures the success, efficiency, and accuracy of task or assessment performance, capturing both completion and quality of outcomes.
- **Usage Behavior Over Time**: Examines engagement across timeframes, focusing on persistence, continuity, and when activity ceases.
- **Technology of Access**: Identifies the technological conditions under which the platform is used, such as device type.

2.3.2.1 Metrics used in health-related studies

Table 2.2: Engagement metrics in health-related studies, with definitions and examples.

Dimension	Metric (definition)	Example	Papers
Time-on-Platform Measures	Time on Task (total time spent on a single defined task)	Time a user spends completing a module	[28]
	Total Time Spent on Platform	Sum of interaction duration with the platform	[28, 50]
	Intensity of Use (average weekly time during active weeks)	Minutes per week during weeks with at least one login	[28]
	Median Screen Display Duration (average screen dwell)	Median time a screen is displayed before switching	[51]

Dimension	Metric (definition)	Example	Papers
Interaction Volume and Frequency	Login Frequency (number of logins)	Total number of logins	[1, 50, 58, 61]
	Number of Entries in Logs (raw interaction count)	Total recorded events	[50]
	Unique Features Used (breadth of functionality accessed)	Number of distinct platform tools accessed	[50, 66]
	Interactions per Task (actions performed within a task)	Number of clicks, inputs, or steps to finish a task	[27, 28, 50]
	Volume (screen views per session)	Average screens visited per session	[51]
Content Access and Navigation	Diversity (unique screens per session)	Mean number of different screens viewed	[51]
	Content Categories Accessed (binary access per category)	Coded 1 if accessed, 0 if not	[51]
Usage Behavior Over Time	Span of Use (temporal extent of use)	Weeks between first and last login	[28]
	Continuous Usage (longest active streak)	Consecutive weeks with at least one login	[28]
	Proportion of Active Weeks (% of active weeks)	Ratio of active to total observed weeks	[28]

The set of health-related studies reveals a smaller but still diverse range of metrics, as shown in Table 2.2. Regarding the time on task metric, the notion of "task" may be interpreted differently from one platform to another, so it is treated here as an activity that can be clearly initiated and completed, thus constituting a measurable unit of activity.

To refine time-based estimates, both Logan [42] and Kadoic & Oreski [36] introduce "user think time", the interval between two consecutive interactions. When this interval exceeds 60 minutes, it is treated as idle, and such periods are excluded, yielding a more accurate measure of active use.

Some metrics are recorded at the session level, such as the "Diversity" one, that measures the mean number of unique screens viewed per session. A session is defined here as a continuous period of active use of the platform, typically beginning with a login or the first activity after a period of inactivity, and ending either with an explicit logout or after a predefined idle timeout.

2.3.2.2 Metrics used in non-health studies

Table 2.3: Engagement metrics in non-health studies, with definitions and examples.

Dimension	Metric (definition)	Example	Papers
Time-on-Platform Measures	Time on Task (total time spent on a single defined task)	Time a user spends completing a module	[20, 23, 25, 40, 42, 54, 69, 70, 73]
	Total Time Spent on Platform	Sum of interaction duration with the platform	[2, 15, 16, 26, 52, 57, 71]
	Daily Average Time	Mean daily minutes logged in	[35]
	Average Time per Page	Avg. seconds spent per page	[2, 9, 15–17]
	Time Spent on Learning Activities (educational content only)	Minutes spent on lectures, exercises, or modules	[24, 62–64]
	Time on Relevant vs. Irrelevant Pages (productive vs. non-productive)	Time spent on pages defined as "learning-relevant" or "irrelevant/orienting"	[31]
	Orientation Time (time from task start to next interaction)	Delay between opening a task and first action	[70]
	Average Time on Quizzes	Minutes per quiz attempt	[15, 16]
	Largest Period of Inactivity (longest idle gap)	Longest streak of days without logging in	[71]
	Punctuality (responsiveness to task availability)	Time taken to start a task after release	[42]
	Late Submissions (missed deadlines)	Tasks submitted past deadline	[71]
Interaction Volume and Frequency	Login Frequency (number of logins)	Average logins per week	[1, 26, 30, 52, 71]

Dimension	Metric (definition)	Example	Papers
	Number of Entries in Logs (raw interaction count)	Total recorded events	[36, 42, 45]
	Total Number of Interactions (overall actions)	All clicks, submissions, or inputs	[20, 40]
	Interactions per Task (action count per task)	Steps taken to complete one assignment	[1, 5, 11, 12, 20, 23, 26, 30, 36, 44, 49, 52, 55, 57, 69]
	Frequency ID (time-of-day usage)	Morning/afternoon/evening/ night login categorization	[52]
	Days Absent (periods without access)	Number of non-active days since first login	[11, 44]
	Class Attendance (being present in classes)	Count of attendances to classes	[30]
Content Access and Navigation	Pages Read	Number of distinct pages read	[9, 15, 16, 24, 62–64]
	Visits per Page (revisitation)	Number of times a page is re-accessed	[2, 17, 23, 71]
	Pages Above/Below Time Threshold (reading depth filter)	Pages considered "read" only if time $>$ threshold	[15]
Task and Performance Outcomes	Task Completion (finished vs. unfinished tasks)	Binary completion measure	[42]
	Attempts per Task (total tries per item)	Number of repeated attempts to solve	[20, 23]
	Activities Passed/Failed (success rate)	Ratio of successful to unsuccessful attempts	[55]
	Questions/Quizzes Completed (engagement in assessments)	Number of quizzes taken	[15, 16, 23, 24, 62]

Dimension	Metric (definition)	Example	Papers
	Correct/Incorrect Answers or Score (accuracy)	% correct responses or overall score	[17, 23, 24, 62]
	Number of Actions per Item (efficiency)	Steps taken to solve a single question	[73]
	Answer Length (response detail in text fields)	Word count of open-ended responses	[54]
	Similarity	Match between user solution path and reference path	[29]
	Efficiency	Efficient vs. inefficient steps compared to standard path	[29]
Usage Behavior Over Time	Day of Last Activity (persistence)	Last recorded login date	[23]
Technology of Access	Technology Used (access channel)	Web vs. mobile access distinction	[60]

2.3.3 Critical Discussion

The reviewed literature demonstrates that user engagement is a multidimensional construct that can be measured in a variety of ways. Across both health-related and non-health-related studies, researchers have proposed metrics regarding time spent on the platform, frequency of interactions, navigation behaviour, content consumption, task completion, and longitudinal usage patterns. However, despite the large number of available measures, no consensus exists regarding which metrics best represent engagement.

The comparison between health-related and non-health-related studies also revealed notable differences. Non-health studies tend to employ a broader and more detailed set of engagement metrics, particularly regarding navigation behaviour, performance outcomes, and learning-related activities. In contrast, health-related studies generally rely on a smaller set of metrics centred around platform usage, activity frequency, and long-term adherence. This difference likely reflects the objectives of the respective domains. Educational and learning platforms often focus on performance and task completion, whereas digital health interventions are more concerned with sustained participation and continued use over time.

Another important observation is that most studies rely heavily on simple activity counts, such as logins, visits, or interaction frequencies. While these measures are easy to compute and interpret, they may not fully capture the quality or meaning of user engagement. A high number of interactions may not necessarily imply meaningful engagement with the intervention, just as low interaction volumes may not necessarily indicate disengagement. Several studies therefore advocate combining multiple dimensions of engagement rather than relying on isolated metrics.

The review also highlights methodological challenges associated with interaction log analysis. Measures based on time spent on the platform are particularly sensitive to assumptions regarding inactivity thresholds and session reconstruction. Different definitions of sessions or active use can produce substantially different estimates of engagement. Similarly, metrics derived from navigation patterns depend on how user actions are recorded and structured within the platform.

Taken together, the literature suggests that engagement cannot be adequately represented by a single metric. Instead, it should be viewed as a multidimensional set.

Chapter 3

Methodology

3.1 Intervention Context

The data used in this dissertation were collected as part of a long-standing collaboration between INESC TEC and Vrije Universiteit Amsterdam (VUA), currently carried out within the RECONNECTED project. Specifically, this work analyzes data from a previous study conducted in the Netherlands, resulting from this collaboration. The description below is included solely to provide context regarding the origin of the dataset, as the contribution of this work begins after the data acquisition. The study was carried out in Dutch and evaluated a four-week internet-based cognitive behavioral therapy (CBT) intervention for adults experiencing low mood, delivered through the Moodbuster platform. Participants were recruited from the general population through social media and an online recruitment service between August 2021 and January 2022. Eligibility criteria included being at least 18 years old, expressing a willingness to improve their mood, and having access to both a computer with an internet connection and a smartphone. Before accessing the intervention, participants completed a set of baseline self-reported assessments, including the Patient Health Questionnaire-9 (PHQ-9), which was used to determine study eligibility and will be described in greater detail later in this chapter. Individuals with moderate to severe depression or at risk of suicide, as identified through this assessment, were excluded and referred to appropriate support services. Eligible participants then received access to the Moodbuster platform and completed a four-week intervention. Following the intervention, they were invited to complete post-intervention assessments, which included measures of depressive symptoms, motivation, usability, and user satisfaction, all of which are described in more detail in the following sections. Participants who completed the post-intervention assessment received a €30 voucher. A total of 106 participants were enrolled in the study.

The intervention was delivered through two complementary components: a web portal and a mobile application. The web portal provided access to a structured therapeutic program consisting of three modules: Introduction, Psychoeducation, and Pleasurable Activities. The Introduction and Psychoeducation modules were available from the beginning of the intervention, while the Pleasurable Activities module became available during the second week. Each module combined educational material with interactive exercises designed to support therapeutic progress.

The mobile application was designed to support daily self-monitoring through ecological momentary assessment, where participants received three semi-random mood-rating prompts per day via notifications. In addition to mood ratings, users could report anxiety and sleep quality levels in this assessments. The application also included a records feature that allowed participants to create free-text entries related to food intake, sleep, and personal thoughts.

Throughout the intervention, interaction data were continuously recorded for both the web portal and the mobile application. The resulting data consist of detailed user interaction records, including navigation, notifications, ratings, records, module progression, and timing information, as well as self-reported assessment measures collected before and after the intervention.

3.2 Data Sources

The analyses in this work draw on two complementary sources of data generated by the prior Moodbuster study: (i) interaction logs from the mobile application and web portal, and (ii) self-reported assessment data collected at baseline and posttreatment. While the interaction logs capture how participants engaged with the intervention in practice, the assessment data provide information on psychological outcomes and participant perceptions of the intervention. Together, these datasets enable the investigation of engagement from both behavioral and self-reported perspectives.

The interaction data consisted of timestamped event-level records stored in structured CSV files. Mobile application logs included lifecycle events (e.g., application opens, closes, and resumes), authentication events, notification delivery and responses, ratings submitted by users, free-text records, and navigation traces. Web portal logs included authentication events, page visits, module progression and completion records, therapist messages, and page-level timing information. For each interaction file, a corresponding `[file]_count.csv` file provided pre-aggregated counts of the same event types.

While the original study enrolled 106 participants, only 101 had recorded platform usage data and were therefore included in the dataset analyzed in this dissertation. Among these, 96 users generated activity on the mobile application and 95 generated activity on the web portal. Six participants contributed only mobile application data, while five contributed only web portal data. Details regarding the self-reported assessment measures are presented later in this chapter.

3.2.1 Application Interaction Data

The mobile application logs participants' daily interactions, including logins and logouts, notifications received, pages and segments visited, as well as ratings and records submitted. Each feature generates its own CSV file:

- **app.csv** App lifecycle events emitted by the client. Records when a user opens the app (`open`), leaves it in the background (`pause`), or brings it back to the foreground (`resume`).

- Fields: `username`, `timestamp`, `event`, `event_date`, `event_time`.
- Example: `U?, 2021-10-01T15:49:09Z, open, 01-10-2021, 15:49:09`

- **app_login.csv** Explicit authentication events into the app. Captures when a user provides credentials and initiates a session, complementing lifecycle events.
  - Fields: `username`, `timestamp`, `date_login_app`, `time_login_app`.
  - Example: `U?, 2021-09-21T20:20:48Z, 21-09-2021, 20:20:48`
- **app_notifications.csv** Notification delivery logs and subsequent user responses. Tracks notification operations: `open` (notification tapped), `delete` (dismissed), or `unknown` (auto-cleared or replaced). The `source` indicates whether the notification originated from the OS (`device`) or the in-app notification center (`app`).
  - Fields: `username`, `timestamp`, `date`, `time`, `notification_id`, `operation`, `source`.
  - Example: `U?, 2021-09-22T13:39:05Z, 22-09-2021, 13:39:05, 1, open, device`
- **app_ratings.csv** In-app rating submissions. The `value` corresponds to an integer representing the rating value. The `source` indicates whether the rating came from a system notification (`device`), an in-app prompt (`app`), or manually (`empty`).
  - Fields: `username`, `timestamp`, `date`, `time`, `value`, `source`.
  - Example: `U?, 2021-09-22T13:39:08Z, 22-09-2021, 13:39:08, 7, device`
- **app_records.csv** User records. Each entry is categorized by `recordtype`, which can be `sleep`, `food` or `thought`, while the `content` field stores feature-specific JSON (including timestamps and free-texts).
  - Fields: `username`, `timestamp`, `date`, `time`, `recordtype`, `content`.
  - Example: `U?, 2021-10-02T12:28:13Z, 02-10-2021, 12:28:13, food, "{""food"":""breakfast"",""date"":""2021-10-02T10:00:58+02:00""}"`
- **app_pages.csv** Page-view telemetry capturing navigation between major app pages.
  - Fields: `username`, `timestamp`, `date`, `time`, `page_visited`.
  - Example: `U?, 2021-10-01T15:49:36Z, 01-10-2021, 15:49:36, Messages`
- **app_segments.csv** Within-page segment views (e.g., tabs, subtabs, sections).
  - Fields: `username`, `timestamp`, `date`, `time`, `segment_visited`, `page`.
  - Example: `U?, 2021-10-01T15:49:36Z, 01-10-2021, 15:49:36, unread, Messages`

3.2.2 Portal Interaction Data

Similarly, the portal records participants' daily interactions, including module completions, activities performed, logins and logouts, page visits with entry and exit times, and messages received from the therapist. Each feature has its own CSV file associated:

- **patient_logs.csv** High-level navigation events on the portal. `operationType` includes page entries and exits (`entryPage`, `exitPage`), as well as logins (`loginW`) and logouts (`logoutW`). `moduleId` identifies the course module; `page` is the within-module page index.
  - Fields: `username`, `date`, `time`, `idLogEntry`, `codPatient`, `serverTimeStamp`, `clientTimeStamp`, `operationType`, `moduleId`, `page`.
  - Example: `U?, 07-02-2022, 09:07:03, 40428, C?, 2022-02-07T09:07:03Z, 2022-02-07T10:04:14Z, entryPage, intro, 1.0`
- **patient_login_logout.csv** Authentication events on the portal. `operationType` can be `loginW`/`logoutW`.
  - Fields: `username`, `date`, `time`, `idLogEntry`, `codPatient`, `serverTimeStamp`, `clientTimeStamp`, `operationType`.
  - Example: `U?, 10-09-2021, 18:15:58, 27539, C?, 2021-09-10T18:15:58Z, 2021-09-10T19:13:25Z, loginW`
- **patient_modules.csv** Module lifecycle and progress summary per user–module: creation, first/last interactions, top/last page and exercise visited, derived time-in-module, and status (e.g., `completed`). Timestamps appear in multiple granularities (full ISO, plus date/time splits).
  - Fields: `username`, `codPatient`, `creationDate`, `creationDateOnly`, `creationTime`, `codModule`, `moduleName`, `maxPagesByModule`, `topPageVisited`, `lastPageVisited`, `maxExercisesInModule`, `topExerciseLog`, `lastExerciseDoneDate`, `lastExerciseDoneDateOnly`, `lastExerciseDoneTime`, `startModuleDate`, `startModuleDateOnly`, `startModuleTime`, `stopModuleDate`, `stopModuleDateOnly`, `stopModuleTime`, `lastOperationDate`, `lastOperationDateOnly`, `lastOperationTime`, `hoursInModule`, `statusId`.
  - Example: `U?, C?, 2021-09-16T21:29:02Z, 16-09-2021, 21:29:02, intro, Introduction, 23, 23, 16, 6, 5, 2021-10-21T08:57:34Z, 21-10-2021, 08:57:34, 2021-09-21T20:24:22Z, 21-09-2021, 20:24:22, 2021-09-23T22:05:47Z, 23-09-2021, 22:05:47, 2021-10-21T08:57:56Z, 21-10-2021, 08:57:56, 51, completed`
- **patient_page_timings.csv** Per-module page timings computed on the portal. Each row records an enter/exit pair and the elapsed time in milliseconds.

- Fields: `username`, `module`, `page`, `enter_time`, `exit_time`, `elapsed_time_milliseconds`.
- Example: `U?, ba, 2, 2021-10-14T09:43:39.917Z, 2021-10-14T09:48:52.977Z, 313060`

- **patient_activities.csv** Information regarding module activities for each patient.
  - Fields: `username`, `title`, `allDay`, `lastUpdateServerTimeStamp`, `activityCod`, `history`, `activityTimestamp`.
  - Example: `U?, Activity, 0, 2022-01-11T10:15:05Z, P, 0, 2022-01-11T10:12:36Z`
- **patient_messages.csv** Therapist messages sent to the patients.
  - Fields: `idMessage`, `receiver`, `title`, `body`, `status`, `received_at`.
  - Example: `9733, U?, Tip, Reminders can help you plan pleasant activities., read, 2021-09-27T13:01:05.387Z`

3.3 Overview of the Processing & Analysis Pipeline

Building on the data sources described above, this work implements a pipeline that combines interaction data from the mobile application and web portal with self-reported assessment data collected at baseline and posttreatment. Figure 3.1 presents a high-level overview of the workflow followed throughout the study.

The first step consists of cleaning the interaction data to ensure data quality and consistency before analysis. Once the data have been prepared, an analysis of the interaction logs is conducted. This stage examines how participants interact with different components of the platform, identifying behavioral patterns, feature usage trends, and common navigation paths. Besides informing the development of engagement metrics, this analysis provides insights into which features are most and least used and highlights potential areas for platform improvement. Such findings can support future efforts to enhance user engagement and, ultimately, the effectiveness of the intervention.

Following this, engagement metrics are constructed from the interaction data. These metrics are then combined with the self-reported assessment measures to support the correlation analyses and predictive modelling tasks presented in this work. The pipeline shown in Figure 3.1 is intentionally presented at a high level, while the specific implementation details of each stage are described in the following sections.

Artificial Intelligence tools were used exclusively as programming assistants during the implementation of this dissertation. Their use was limited to assisting in the development of the code used to generate the visualizations for the interaction data analysis and to producing an initial implementation skeleton for the correlation analysis and predictive modelling pipelines. All generated code was critically reviewed, modified where necessary, and tested before being incorporated into the final implementation. No AI tools were used to interpret the results or formulate the conclusions of this work.

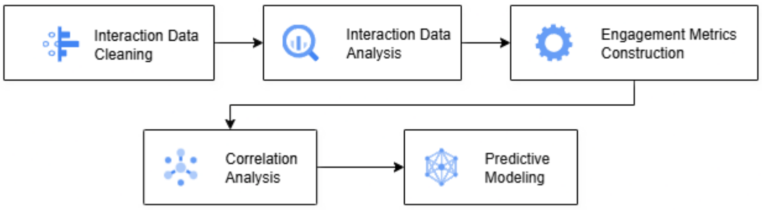


Figure 3.1: Overview of the pipeline workflow.

3.4 Interaction Data Cleaning

This section describes the procedures applied to prepare the mobile application and web portal interaction logs for analysis. The objective was to ensure data consistency, identify potential quality issues, and detect anomalous usage patterns before conducting the interaction analyses and constructing engagement metrics. Figure 3.2 presents an overview of this process, that included whitespace stripping to every column in the interaction data files, inspection of data completeness and frequency distributions, checking and removable of duplicate records, enforcement of structural consistency constraints, and examination of outliers across general logs.

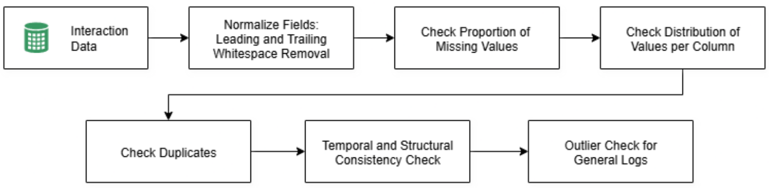


Figure 3.2: Overview of the interaction data cleaning steps.

3.4.1 Data Completeness and Value Distributions

Data completeness and value distributions were examined for each interaction file. The proportion of missing values was calculated for every column containing at least one missing entry. In addition, the frequency distribution of values in each column was inspected. Columns in which a single value accounted for more than 50% of all observations were flagged for further inspection. This threshold was used to identify potential constants, low-variability attributes, or possible logging issues that could affect later analyses.

3.4.2 Duplicates and Structural Consistency

Beyond completeness and concentration, the interaction data were examined for duplicate entries and structural inconsistencies to preserve temporal and quantitative integrity before downstream analyses. Consistency checks were carried out where they were meaningful; in practice, all applied to patient_modules.csv and are listed in Table 3.1. These include verifying that completion timestamps follow starts, that reported page counts match navigation records, and that the last exercise done belongs to the corresponding module.

Table 3.1: Consistency constraints applied to patient_modules.csv.

Constraint	**Description**
startModuleDate $\leq$ stopModuleDate	A module's start date must not be later than its stop date.
startModuleDate $\leq$ lastExerciseDoneDate	A module's start date must precede or coincide with the last recorded exercise.
startModuleDate $\leq$ lastOperationDate	A module's start date must precede or coincide with the last operation performed.
creationDate $\leq$ startModuleDate	A module cannot start before its creation timestamp.
creationDate $\leq$ stopModuleDate	A module cannot be stopped before its creation.
creationDate $\leq$ lastExerciseDoneDate	Exercises cannot be completed before the module is created.
creationDate $\leq$ lastOperationDate	Operations cannot occur before the module exists.
maxPagesByModule $\geq$ topPageVisited	The highest visited page must not exceed the maximum defined pages of the module.
maxPagesByModule $\geq$ lastPageVisited	The last visited page must not exceed the module's maximum defined pages.
topPageVisited $\geq$ lastPageVisited	The index of the top (highest) page visited must not be lower than the last visited page.

Continued on next page

Constraint	Description
`maxExercisesInModule` $\geq$ `topExerciseLog`	The maximum defined number of exercises must not be exceeded by the top logged exercise.

3.4.3 Outlier screening

The general app and portal logs were also screened for outliers to identify unusual usage patterns that could indicate logging errors or non-human activity. For each user, we checked the total number of recorded events and the time gaps between consecutive events, and inspected their distributions. Users with extreme values were identified using a 3×Interquartile Range (IQR) criterion. Each identified outlier was then manually reviewed to determine whether it could be attributed to logging errors or non-human activity, rather than genuine user behavior. The outcome of this analysis is presented in the Results chapter.

3.5 Interaction Data Analysis

This section specifies the analysis conducted on the interaction data to characterize platform usage and guide the construction of engagement metrics. These analyses examined behavioral patterns, temporal dynamics, and feature-level usage trends, with the aim of providing insights to improve the platform and support its effectiveness.

3.5.1 App Notifications

Patients can receive notifications from the mobile app or directly from the device. These notifications are intended to prompt users to return to the app and complete their daily mood ratings. Several aspects of notification performance were examined:

- **Volume and response rates:** overall and source-specific (app vs. device) notification counts and open rates.
- **Temporal patterns:** distribution of notification opens across hours of the day and days of the week, to identify periods of higher user responsiveness.
- **Notification to app engagement:** whether a notification was followed by an app open or resume event within 60 minutes, and the distribution of time elapsed between the notification and the subsequent app event.
- **Notification to rating conversion:** whether a notification was followed by a rating submission within 60 minutes, used to assess the extent to which ratings are driven by notifications rather than initiated spontaneously.

- **Engagement fatigue:** relationship between the volume of notifications received and individual open rates, to assess whether higher notification volume is associated with declining responsiveness.

3.5.2 App Ratings

Ratings are the main feature of the mobile application and can be submitted following a notification from the app, from the device, or manually without any notification prompt. Rating behavior was analyzed in terms of its overall volume and distribution across these different sources.

3.5.3 App Records

Records are another feature of the mobile application, consisting of open-text fields where patients could write information about their food intake, sleep, and ongoing thoughts. Record data were analyzed to characterize their volume, textual properties, and associations with sleep quality. The following analyses were conducted:

- **Volume by type:** total and weekly counts per record type, as well as their distribution across weekdays.
- **Text characteristics:** character length of entries for record types containing free-text responses (Food and Thought).
- **Sleep durations:** distributional summaries of reported sleep durations.

3.5.4 App Navigation

User navigation within the app was analyzed to characterize session structure, page usage, and flow patterns. The following analyses were conducted:

- **Page level analysis:** most visited pages and median dwell time per page
- **Session structure:** session duration distribution; total and unique pages visited per session; user activity streaks measured as the longest consecutive sequence of active days; and maximum inactivity gaps measured as the longest consecutive sequence of inactive days.
- **Transitions and flows:** page-to-page transition analysis and Sankey diagrams for the most frequent flows
- **Sequence mining:** frequent navigation sequences.

3.5.5 Portal Logins

User authentication data from the portal was analyzed to characterize patterns of access and engagement. Daily login activity was examined both in terms of overall frequency and its temporal distribution. Specifically, the following analysis was conducted:

- **Daily logins and intensity:** totals over time and share of logins by weekday/hour.

3.5.6 Portal Modules

Learning module progression was analyzed to evaluate how users advanced through the assigned therapeutic content, the time required to complete each module, and points at which engagement declined. The analyses considered overall completion rates, pacing, content coverage, and potential bottlenecks in the progression flow. The following aspects were examined:

- **Completion funnel:** for each module, the proportion of users who started, reached intermediate progression milestones, and ultimately completed the module.
- **Throughput time:** time elapsed between a user's first event in a module and its completion, summarized by median and interquartile range per module.
- **Coverage:** percentage of pages visited and exercises completed per module per user.
- **Bottlenecks:** scatter plots of completion time against completion status to highlight slow or non-completing cases; slowest pages by median dwell time; pages with the highest revisit rates; and pages revisited across different days, indicating content that required multiple sessions to process.
- **Stalling:** longest period of inactivity between successive modules in a user's assigned progression path.

3.5.7 Portal Messages

Therapist messages exchanged through the web portal were analyzed to assess whether they acted as triggers for subsequent user engagement and progression through the intervention. To distinguish message-related activity from users' normal behavior, post-message activity rates were compared with baseline activity observed during equivalent periods preceding message delivery.

The following outcomes were examined:

- **Message to login conversion:** proportion of messages followed by a portal login within 24 hours, compared with the baseline login rate during the 24 hours preceding message delivery.
- **Message to module activity conversion:** proportion of messages followed by any interaction with module content within 48 hours, compared with baseline module activity during the preceding 48 hours.
- **Message to progression events:** proportion of messages followed by module start events and module completion events within seven days, compared with corresponding baseline rates.

3.6 Engagement Metrics Computation

This section presents the engagement metrics for the web portal and the mobile app. The metrics were selected carefully, influenced by findings from the literature review and by the constraints and opportunities of the available data. They seek to obtain a complete understanding of user engagement, covering everything from basic activity metrics to usage trends over time, behavioral tendencies, and interactions specific to features. Particular attention was given to constructing sessions, estimating time-on-page, and login-based behaviors, considering the available interaction data.

3.6.1 App Metrics

For the app, engagement metrics were defined with respect to user sessions, page visits, notifications, ratings, and records. Sessions are delimited by `open` and `resume` events and end with a `pause`. If more than 30 minutes pass after an `open` or `resume`, the period is treated as inactivity and no session is counted. The time spent on pages is calculated by examining each `page_visited` event: if the following event occurs within 30 minutes, the time interval is attributed to the page, also if the next event is a `pause` within 30 minutes, the time difference is considered too, whereas if neither condition applies, the duration is ignored to avoid inflating dwell time. Temporal preference is represented as an ID in $\{1,2,3,4\}$, with 1 indicating major use from 06:00–11:59, 2 from 12:00–17:59, 3 from 18:00–21:59, and 4 for all other times. Table 3.2 shows the full range of app engagement metrics along with their definition. Detailed computation rules for each metric are provided in Appendix A.1.

Table 3.2: App engagement metrics by dimension with definitions.

Dimension	Metric	Definition
Time-on-Platform Measures	Total time spent on the app	**Definition:** Sum of session durations per user.
	Average time spent in pages	**Definition:** Mean per-user page dwell time. For each page visit, dwell time is the interval from that page visit to the immediately following event (next page visit or a pause) occurring within 30 minutes.

Dimension	**Metric**	**Definition**
	Average time spent in active weeks	**Definition:** Mean per-user weekly usage time, computed only across **active weeks**. An active week is a week in which the user has at least one valid start–end interval (an `app.event` $\in$ {"open","resume"} followed by the next `app.event = "pause"`) whose duration is $0 < \Delta t \leq 30$ minutes.
	Average duration of a session	**Definition:** Mean per-user duration of sessions, in minutes.
	Average daily time	**Definition:** Mean per-user **daily** usage time. Days on which the user has at least one valid start–end interval (an `app.event` $\in$ {"open","resume"} followed by the next `app.event = "pause"`) with duration $0 < \Delta t \leq 30$ minutes.
Interaction Volume and Frequency	App opens / pauses / resumes	**App opens — Definition:** Total number of **open** events per user. **App pauses (closes) — Definition:** Total number of **pause** events per user. **App resumes — Definition:** Total number of **resume** events per user.
	Daily average number of visits	**Definition:** Mean per-user number of visits per day, where a visit is proxied by `app.event` $\in$ {"open","resume"}. Averaged **across days that have at least one open/resume**.
	Total number of entries in the logs	**Definition:** Total count of log entries per user across all app-related files (events, pages, ratings, records, notifications, segments, logins).
	Number of ratings	**Definition:** Total number of rating actions performed by the user.
	Number of records	**Definition:** Counts of submitted records per type (Food, Thought, Sleep) for each user.

Dimension	Metric	Definition
	Days of absence since first login	**Definition:** Number of days between the user's **first login date** and **last observed day** (based on a pause) on which the user had **no access events** (no `app.event` $\in$ {"open", "resume"}).
	Temporal preference	**Definition:** Dominant time-of-day period in which the user generates the **most events**, mapped to an ID: `1=morning (06:00-11:59)`, `2=afternoon (12:00-17:59)`, `3=evening (18:00-21:59)`, `4=night (otherwise)`.
Content Access and Navigation	Average pages per session	**Definition:** Mean number of page visits per session for each user.
	Average unique pages per session	**Definition:** Mean number of **distinct** pages visited per session for each user.
Task and Performance Outcomes	% ratings without notification	**Definition:** Share of ratings submitted **without a notification trigger**, i.e., rows where `app_ratings.source = ""` (empty after trimming).
	% ratings via OS notification	**Definition:** Share of ratings submitted **after an OS/device notification**, i.e., rows where `app_ratings.source = "device"`.
	Average length of record answers	**Definition:** Mean **character length** of text responses in `Food` and `Thought` records, computed per user and per record type.
Usage Behavior Over Time	Longest consecutive stretch of active weekly usage	**Definition:** Length (in ISO weeks) of the **longest run of consecutive weeks** during which the user had **at least one access event** (`app.event` $\in$ {"open", "resume"}) within the observation window $[\text{first login date}, \text{last observed date}]$. The **last observed date** is taken as `date(max app.timestamp)` where `app.event = "pause"`.

Dimension	Metric	Definition
	% of weeks active after first login	**Definition:** Percentage of ISO weeks in the user's observation window $[\text{first login date}, \text{last observed date}]$ that contain **at least one access event** (`app.event` $\in \{\text{“open”}, \text{“resume”}\}$).

3.6.2 Portal Metrics

For the portal, engagement metrics focus on session durations, login activity, and module progress. In contrast to the app, the portal's login/logout system isn't completely reliable as many users do not intentionally log out but rather close the browser. To account for this, session durations are derived by checking each `loginW` event and determining whether it is followed by a `logoutW` or by another `loginW`. In the former case, the session duration is the difference between the two events. In the latter, the duration is calculated as the difference between the `loginW` and the last recorded interaction before the next `loginW`. Similarly to the app, a temporal preference ID ($\{1,2,3,4\}$) identifies the main time of day when the user is engaged. Table 3.3 presents the complete collection of portal engagement metrics with their definition and computation. Detailed computation rules for each metric are provided in Appendix A.1.

Table 3.3: Portal engagement metrics by dimension, with definitions.

Dimension	Metric	Definition
Time-on-Platform Measures	Total time spent on the portal	**Definition:** Sum of all **login sessions** per user, in minutes. A session starts at `patient_logs.operationType = "loginW"` and ends at the next `"logoutW"` for the same `patient_logs.username`; if a new `"loginW"` occurs before a `"logoutW"`, the previous session is closed at the `last observed event time` for that user.
	Total time spent on modules	**Definition:** Sum of measured time spent within **module pages**, in minutes, per user.
	Average time spent in pages	**Definition:** Mean per-user **page dwell time** on the portal, in minutes.

Dimension	Metric	Definition
	Average time spent in active weeks	**Definition:** Mean per-user **weekly usage time**, computed **only across active weeks**. An **active week** is a calendar week that accumulates **>0 seconds** of reconstructed session time.
	Average duration of a session	**Definition:** Mean per-user duration of **login sessions**, in minutes.
	Average daily time	**Definition:** Mean per-user daily usage time, in minutes, averaged **over active days only**. An **active day** is any calendar day on which at least one reconstructed session contributes > 0 seconds.
	Average time to complete a module	**Definition:** Mean per-user time spent to complete a module.
Interaction Volume and Frequency	Number of logins	**Definition:** Total count of login events per user.
	Daily average number of visits	**Definition:** Mean number of **visits** per **access day**, where an access day is any day with at least one `loginW`.
	Total number of entries in the logs	**Definition:** Total count of log entries per user across `patient_logs.csv` (all operation types).
Content Access and Navigation	Average module pages per session	**Definition:** Mean number of **module pages** visited per login session, per user.
	Average unique module pages per session	**Definition:** Mean number of **distinct module pages** visited per session, per user. Distinctness is computed on the pair `(module, page)` within each session window.
Task and Performance Outcomes	Number of modules completed	**Definition:** Total count of modules whose status is `"completed"` for each user.
	% of modules completed	**Definition:** Share of a user's modules that are `"completed"` among all modules.

Dimension	**Metric**	**Definition**
Usage Behavior Over Time	Days of absence since first login	**Definition:** Number of **calendar days without any login** between the user's first and last login dates (inclusive). A "login day" is any date with `operationType = "loginW"`.
	Longest consecutive stretch of active weekly usage	**Definition:** Number of consecutive weeks during which the user logged in at least once.
	% of weeks active after first login	**Definition:** Percentage of weeks between the user's first and last login weeks (inclusive).
	Temporal preference	**Definition:** Dominant time-of-day period (by login count) mapped to an ID: `1=morning (06:00-11:59)`, `2=afternoon (12:00-17:59)`, `3=evening (18:00-21:59)`, `4=night (otherwise)`.

3.6.3 Metric Post-Processing

All app and portal-level metrics were consolidated into a single user-level dataset containing one row per user. After merging, per-user missingness was checked across all metric columns to identify which metrics were unavailable for each user.

Since many users never submitted records, record-related metrics were frequently missing. Users with no records had missing values for record count metrics and average record text length metrics. Since these are activity-based measures, the absence of entries was interpreted as absence of activity and imputed as zero. Likewise, for users without any observed module viewing or completion, module-related metrics were set to zero. Coverage asymmetries were also documented, as some users contributed only app data while others contributed only portal data. This distinction was explicitly recorded and taken into account in subsequent analyses.

In addition, an outlier screening was performed for each metric independently using a 3×IQR criterion. Users identified as outliers on more than two metrics were flagged separately for the mobile application and the web portal. These flagged users were retained in the dataset and carried forward to later analytical steps.

Overall, most users were not identified as outliers on any engagement metric: in the mobile application dataset, 8 users were identified as outliers on two or more metrics, while in the portal dataset, although outliers appeared more frequently at the individual metric level, only 13 users were identified on two or more metrics overall. Users identified as outliers were retained and used in later analyses by repeating selected analyses both including and excluding them, in order to evaluate whether conclusions were sensitive to their presence.

3.7 Self-reported Assessment Data

Alongside the interaction logs generated by the mobile application and web portal, participants in the intervention completed self-reported assessments at baseline and posttreatment. These assessments provide information about participants' motivation, depressive symptom severity, perceptions of the intervention, and adherence to the study protocol. While the interaction logs capture objective behavioral engagement, the assessment data offer a complementary perspective on participants' experiences and outcomes, allowing engagement measures to be examined in relation to therapeutic change.

Several standardized instruments were included in the dataset. Perceived usability was measured at posttreatment using the System Usability Scale (SUS), a 10-item questionnaire scored on a five-point Likert scale. Responses are converted into a total score ranging from 0 to 100, with higher values indicating greater perceived usability of the intervention.

Participant satisfaction was assessed using the Client Satisfaction Questionnaire for Internet-based Interventionss (CSQ-Is), also administered at posttreatment. The instrument consists of eight items scored from 1 to 4, producing a total score between 8 and 32, where higher scores reflect greater satisfaction with the intervention.

Depressive symptom severity was measured using two different instruments. As previously described, the Patient Health Questionnaire-9 (PHQ-9) was administered as part of the baseline assessment and served both as a measure of baseline depressive symptom severity and as a screening instrument for participant eligibility. The PHQ-9 is a nine-item questionnaire with item scores ranging from 0 to 3 and a total score between 0 and 27. Participants scoring 15 or higher, indicating moderate to severe depressive symptoms, as well as those reporting suicidal ideation, were excluded from the study. To evaluate changes in depressive symptoms throughout the intervention, the Depression subscale of the Hospital Anxiety and Depression Scale (HADS-D) was administered at both baseline and post-treatment. The HADS-D consists of seven items scored from 0 to 3, resulting in a total score ranging from 0 to 21, where higher scores indicate greater symptom severity.

Motivation for treatment was measured using the Short Motivation Feedback List (SMFL), which was also administered at baseline and posttreatment. The instrument contains eight items rated on a scale from 0 to 10, yielding a total score between 0 and 80, with higher scores representing greater treatment motivation.

In addition to these questionnaire measures, two adherence indicators were available. The first recorded whether participants continued using the Moodbuster intervention until the end of the four-week study period. The second indicated whether participants completed both the intervention and the posttreatment assessment.

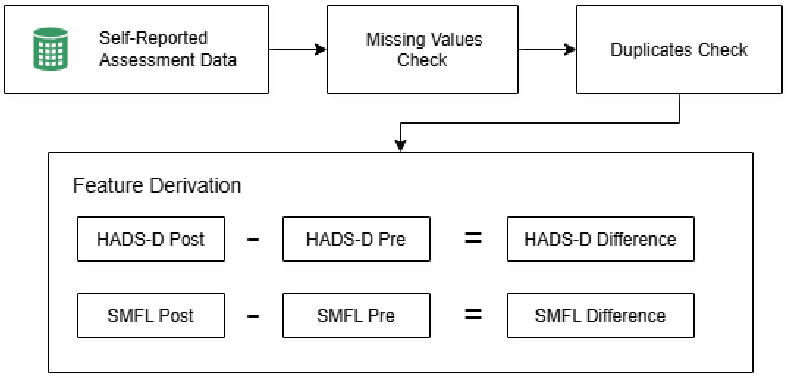


Figure 3.3: Overview of the self-reported assessment data cleaning and feature derivation steps.

3.7.1 Data Cleaning and Feature Derivation

At the time of data acquisition, self-reported assessment data were available for 92 participants. The procedure applied to this dataset is summarized in Figure 3.3.

The data were first inspected for missing values and duplicate records. Missingness was evaluated at the participant level: most participants had complete or near-complete evaluation records, although 9 users were missing the majority of their assessment data and could not reliably contribute to subsequent correlation or predictive analyses. Any inconsistencies identified during this stage were resolved before further processing. Since both the SMFL and HADS-D were administered at baseline and posttreatment, difference scores were computed to capture changes occurring during the intervention period. For each participant, the baseline score was subtracted from the posttreatment score, producing measures of change in treatment motivation and depressive symptom severity.

Following the derivation of these outcome variables, the assessment dataset was matched with the engagement metric dataset using participant identifiers. Only users with information available in both sources were retained for the subsequent correlation and predictive modelling analyses. After this matching process, 9 participants were excluded due to incomplete assessment data, resulting in a final sample of 83 users.

3.8 Correlation Analysis

To investigate the relationship between user engagement and intervention outcomes, correlation analyses were performed between the engagement metrics derived from the interaction logs and the self-reported assessment measures collected during the study. Because engagement data are often highly skewed and may not satisfy the assumptions required for parametric correlation methods, Spearman's rank correlation coefficient (ρ) was used throughout the analysis. Unlike Pearson's correlation coefficient, Spearman's ρ evaluates monotonic relationships based on ranked values, making it more suitable for behavioral data. The coefficient ρ itself functions as a measure of effect size, indicating the strength and direction of an association on a scale from -1 to 1. This was complemented by a significance test for each correlation, with the resulting p-value indicating whether the observed association was statistically significant.

Two complementary correlation analyses were conducted. The first explored the relationships between the self-reported assessment measures themselves. In particular, perceived usability (SUS), user satisfaction (CSQ-I), baseline depression severity (PHQ-9), and the two adherence indicators were correlated with changes in motivation (SMFL difference) and depressive symptoms (HADS-D difference). This analysis aimed to determine whether participants who reported higher usability, satisfaction, adherence, or baseline symptom severity tended to experience different outcomes.

The second analysis examined the relationship between engagement metrics and the differences in motivation and depressive symptoms. Each engagement metric was correlated with both SMFL difference and HADS-D difference to identify which patterns of platform use were most strongly associated with changes in motivation and depressive symptom severity over the course of the intervention.

Since the mobile application and web portal serve different purposes within the intervention and generate distinct types of interaction data, correlations were computed separately for each platform. Furthermore, only participants with recorded activity on a given platform were included in the corresponding analysis to ensure that engagement measures reflected actual usage behavior.

To evaluate the robustness of the findings, the engagement metric correlations were initially calculated under two conditions: using the complete sample, and after removing users previously identified as outliers in more than two engagement metrics. This comparison was used to assess whether observed associations were driven by a small number of atypical users rather than reflecting genuine patterns of engagement. As the two sets of results were broadly consistent, the outlier-removed sample was adopted as the primary analysis reported in this section, since it provides a more representative picture of typical engagement-outcome associations without the influence of atypical users.

Given the large number of pairwise correlations computed across engagement metrics and outcome measures, some associations were expected to meet conventional significance thresholds ($p < 0.05$) by chance alone. For this reason, p-values are reported alongside each correlation coefficient to provide additional context, but they are interpreted in an exploratory rather than confirmatory manner. Consequently, the findings are discussed primarily in terms of the strength and direction of the associations, as reflected by Spearman's ρ, rather than on whether a particular correlation reached a predefined significance threshold.

3.8.1 Temporal Preference Analysis

One engagement metric required a separate analytical approach. The temporal preference indicator represents the period of the day in which a participant most frequently interacted with the platform, categorizing users into morning, afternoon, evening, or night groups. Because these categories are nominal and do not possess a meaningful numerical ordering, they are not appropriate for standard correlation analysis.

To explore whether temporal usage preferences were associated with intervention outcomes, distributions of SMFL difference and HADS-D difference were compared across the four temporal groups using boxplots.

As with the main correlation analyses, temporal preference was examined separately for the mobile application and the web portal, since usage patterns may differ substantially between the two platforms.

3.9 Predictive Modelling

In addition to the correlation analysis, a predictive modelling approach was employed to investigate whether engagement metrics could be used collectively to identify users who experienced improvements in depressive symptoms during the intervention. Whereas correlation analysis considers each engagement metric independently, predictive modelling evaluates all metrics simultaneously, allowing the identification of combinations of behaviors that may be associated with therapeutic outcomes.

The target variable for prediction was derived from the HADS-D assessment. For each participant, a change score was calculated by subtracting the baseline HADS-D score from the posttreatment score. This value was subsequently transformed into a binary outcome. Participants who showed any reduction in depressive symptoms were assigned to the *improved* class, while those whose symptoms remained unchanged or increased were assigned to the *not improved* class.

The predictor variables consisted of the engagement metrics derived from the interaction logs. Models were developed separately for the mobile application and the web portal to account for the distinct nature of engagement on each platform. Only participants with recorded activity on the respective platform were included in each model. The temporal preference metric was excluded from this analysis because it represents a categorical grouping rather than a continuous behavioral measure and therefore is not directly comparable to the remaining engagement indicators.

Logistic regression with L2 regularisation was selected as the modelling technique. This method is widely used for binary classification problems and offers the additional advantage of producing interpretable coefficients for each predictor. These coefficients indicate whether higher values of a given engagement metric are associated with a greater or lower likelihood of symptom improvement.

To ensure comparability between metrics measured on different scales, all predictors were standardised prior to model training. This means each engagement metric was converted to a common scale, so that values originally expressed in different units, such as minutes, counts, or percentages, could be compared fairly. Without this step, metrics with naturally larger numbers could disproportionately influence the model compared to metrics with smaller numbers, even if both were equally important.

In practice, standardisation was implemented using the StandardScaler function from Python's scikit-learn library, which was included as part of the modelling pipeline directly before the logistic regression classifier. During cross-validation, the scaling parameters were calculated separately for each training subset and then applied to the corresponding test subset. This ensured that information from the test data was not used when preparing the training data, thereby preventing data leakage and providing a more reliable estimate of model performance.

Given the relatively limited sample size, model performance was evaluated using stratified five-fold cross-validation. In this procedure, participants were divided into five subsets while maintaining the proportion of improved and non-improved users within each fold. The model was then trained on four folds and evaluated on the remaining fold, repeating the process until each fold had served once as the test set. Performance estimates were obtained by averaging the results across all folds.

Three evaluation metrics were used to assess model performance: accuracy, F_1 score, and AUC-ROC. Accuracy measures the proportion of correctly classified cases, while the F_1 score provides a balance between precision and recall for the improved class. The AUC-ROC evaluates the model's ability to distinguish between improved and non-improved users across different classification thresholds.

To complement the cross-validation results, the final logistic regression model was also fitted using the full dataset for each platform. The resulting coefficients were examined to identify which engagement metrics contributed most strongly to the prediction of symptom improvement. Positive coefficients indicate that higher values of a metric increase the likelihood of belonging to the improved class, whereas negative coefficients indicate the opposite.

Chapter 4

Results

This chapter presents the outcomes of the analyses described in Chapter 3. While the previous chapter focused on data preparation, metric construction, and analytical procedures, the present one reports the empirical findings. Results are organized to follow the same structure as the methods: beginning with interaction data cleaning summary, moving through the analyses of the interaction data, followed by the correlation analysis results, and finally the predictive modeling outcomes.

4.1 Interaction Data Cleaning

The interaction data were examined for missingness, dominant value distributions, duplicate records, and structural inconsistencies. The findings and decisions resulting from each of these checks are described in this section.

4.1.1 Missingness and frequency distribution analysis

Before inspecting individual files, a variable-level screening step was performed to determine which columns of the raw data would progress for the following steps. Rather than removing individual cases, this step assessed each variable as a whole and decided whether to retain or discard it from subsequent steps, guided by two general principles: whether the variable contained sufficient variability to be analytically useful, and whether it was relevant for the construction of engagement metrics. Variables that were entirely missing or fully constant were removed, as they do not provide any information for subsequent analyses. Variables with substantial missingness were retained when the available observations captured important aspects of user behavior, with missing values handled appropriately during engagement metric construction. Similarly, variables dominated by a single value were evaluated on a case-by-case basis. Such variables were retained when the minority categories still conveyed meaningful distinctions in user behavior, and removed otherwise. In some situations, missingness or highly skewed distributions were considered expected given the intervention design. For example, fields related to module completion naturally remained empty for participants who never completed the corresponding module. Decisions taken regarding this step are explained later in the section.

Table 4.1 presents a summary of missingness across all variables.

Table 4.1: Summary of missingness identified in each column of interaction logs.

File	**Column**	**Missing Count**	**Missing %**
app_notifications.csv	operation	755	10.82
app_notifications.csv	source	755	10.82
app_ratings.csv	source	897	18.15
app.csv	event	5708	30.86
patient_activities.csv	description	1541	100.00
patient_activities.csv	url	1541	100.00
patient_activities.csv	reminderMinutesBefore	1541	100.00
patient_activities.csv	questionNumber	1145	74.30
patient_activities.csv	activityDescription	1541	100.00
patient_activities.csv	ImportantActivityTimestamp	1541	100.00
patient_activities.csv	exerciseType	1541	100.00
patient_activities.csv	exerciseText	1541	100.00
patient_activities.csv	exerciseHistory	1541	100.00
patient_activities.csv	exerciseTimestamp	1541	100.00
patient_login_logout.csv	clientTimeStamp	9	0.94
patient_login_logout.csv	exerciseAndStep	957	100.00
patient_login_logout.csv	moduleId	957	100.00
patient_login_logout.csv	page	957	100.00
patient_logs.csv	clientTimeStamp	9	0.07
patient_logs.csv	exerciseAndStep	9591	70.28
patient_logs.csv	moduleId	957	7.01
patient_logs.csv	page	957	7.01
patient_messages.csv	read_at	6795	95.92
patient_modules.csv	topPageVisited	8	3.21
patient_modules.csv	lastPageVisited	8	3.21

File	Column	Missing Count	Missing %
patient_modules.csv	topExerciseLog	18	7.23
patient_modules.csv	lastExerciseDoneDate	18	7.23
patient_modules.csv	lastExerciseDoneDateOnly	18	7.23
patient_modules.csv	lastExerciseDoneTime	18	7.23
patient_modules.csv	stopModuleDate	53	21.29
patient_modules.csv	stopModuleDateOnly	53	21.29
patient_modules.csv	stopModuleTime	53	21.29
patient_modules.csv	lastOperationDate	8	3.21
patient_modules.csv	lastOperationDateOnly	8	3.21
patient_modules.csv	lastOperationTime	8	3.21
patient_modules.csv	hoursInModule	53	21.29
patient_page_timings.csv	exit_time	2640	35.50
patient_page_timings.csv	elapsed_time_milliseconds	2640	35.50

Table 4.2 presents a summary of the frequency distribution of values across the variables for which a single value accounted for more than 50% of all observations.

Table 4.2: Summary of frequency distribution of values for each column in interaction logs.

File	Column	Most Common Value	Count	Percent %
app_notifications.csv	operation	open	4117	66.14
app_notifications.csv	source	app	3225	51.81
app_pages.csv	page_visited	Rate	5141	53.07
app_ratings.csv	type	mood	4941	100.00
app_ratings.csv	source	device	2983	73.76
app_segments.csv	page	Rate	10493	84.40
patient_activities.csv	allDay	0	1399	90.79
patient_activities.csv	periodic	0	1541	100.00
patient_activities.csv	activityCod	P	1037	67.29

File	Column	Most Common Value	Count	Percent %
patient_activities.csv	history	0	1208	78.39
patient_login_logout.csv	systemType	web	957	100.00
patient_login_logout.csv	operationType	loginW	839	87.67
patient_logs.csv	systemType	web	13647	100.00
patient_logs.csv	operationType	entryPage	7222	52.92
patient_logs.csv	moduleId	intro	6360	50.12
patient_messages.csv	sender	therapeut	7084	100.00
patient_messages.csv	title	Goed gedaan!	3879	54.76
patient_messages.csv	status	sent	5846	82.52
patient_modules.csv	statusId	completed	196	78.71
patient_page_timings.csv	module	intro	3767	50.65
patient_state.csv	treatmentState	active	106	100.00
patient_state.csv	languageId	NL	106	100.00
patient_state.csv	onSelfTreatment	1	71	66.98

The inspection of missingness and frequency distributions showed varied patterns among the different files as seen in Tables 4.1 and 4.2. Some variables contained only minor irregularities, while others were either entirely empty or carried negligible information content. The decisions applied to each case are detailed below.

App-related logs. In `app_notifications.csv`, the variables *operation* and *source* showed approximately 11% missingness. Given their relevance for characterizing notification behavior, they were retained. The same applied to `app_ratings.csv`, where the column *source* exhibited 18% missingness but remained conceptually important. By contrast, the *type* field in the same file was constant (*mood* for 100% of cases), and thus removed as uninformative. In `app.csv`, the column *event* was missing in 30.9% of rows. Although this represents a substantial proportion, the rows were retained in order to preserve complete information about each user login. The missingness was handled during the construction of engagement metrics, ensuring that the absence of event labels did not distort subsequent calculations. In `app_pages.csv`, the *page_visited* variable was retained despite moderate skew: while "Rate" accounted for 53% of visits, other page types (e.g., *Notifications*, *Messages*) provided meaningful information on secondary navigation paths. Similarly, in `app_segments.csv`, although “Rate” dominated 84% of entries, the variable was preserved to capture remaining heterogeneity.

Patient activities. Several columns in `patient_activities.csv` (*description*, *url*, *activityDescription*, *exerciseType*, *ImportantActivityTimestamp*) were 100% missing and therefore removed. Field *questionNumber* has 74% of the values missing and doesn’t add any interpretive value so it was excluded. Among similarity checks, *periodic* was entirely constant, so it was excluded. Others were highly skewed but still retained because minority values held interpretive value.

Patient login/logout logs. In `patient_login_logout.csv`, *clientTimeStamp* showed negligible missingness (<1%) and was retained. In contrast, variables such as *exerciseAndStep*, *moduleId*, and *page* were entirely missing and thus excluded. The *systemType* field was constant (“web” for all cases) and removed, while *operationType* was highly skewed (88% “loginW”) but retained due to the relevance of distinguishing login from rarer logout events.

Patient logs. Within `patient_logs.csv`, *clientTimeStamp* again showed only minimal missingness (0.07%) and was retained. The variable *exerciseAndStep* had 70% of the values missing, and didn’t add any interpretive value, so it was excluded. *moduleId* (7% missing but largely skewed toward “intro”) was kept since it provided information about the visited module, making it very valuable. *SystemType* was constant and removed. Conversely, *operationType* (53% “entryPage”) was preserved, as it differentiated navigation behaviors despite skew.

Patient messages. The field *read_at* was 96% missing and therefore excluded. Other message fields revealed very low informational diversity: *sender* was constant (“therapeut”) and removed, while *title* (55% “Goed gedaan!”) and *status* (83% “sent”) were retained, as the minority categories conveyed meaningful distinctions in therapeutic communication.

Patient modules. Most variables in `patient_modules.csv` exhibited low missingness (<8%) and were retained. A few fields, such as *stopModuleDate* and related temporal variables, showed around 21% missingness. This was considered expected, since users who never completed certain modules naturally lacked values in these fields. For the same reason, these variables were retained, as the missingness itself reflects meaningful user behavior. Likewise, the field *statusId* showed high similarity (79% “completed”), but this pattern was interpreted as expected given that the majority of participants did complete their modules. Therefore, all variables in this file were preserved for subsequent analyses.

Patient page timings. In `patient_page_timings.csv`, *exit_time* and *elapsed_time_milliseconds* were missing for roughly 36% of rows. These were kept because timing information is crucial for characterizing engagement. However, within the similarity analysis, *module* was dominated by "intro" (51%) but still retained, as that module was the most popular and the total number of modules was very low.

Patient state. The file `patient_state.csv` contained two fields that were fully constant: *treatmentState* ("active") and *languageId* ("NL"). These were excluded as they did not provide any variability or analytical value. By contrast, the variable *onSelfTreatment* showed 67% of cases equal to "1", but since this field is binary by nature, such skew was considered expected. It was therefore retained.

4.1.2 Duplicates and Inconsistencies

Duplicate records and structural inconsistencies were examined to ensure the temporal and logical coherence of the interaction logs before subsequent analyses.

A small number of duplicates were identified: specifically, four duplicate rows were found in `patient_page_timings.csv`. In each case, only one instance of the duplicated record was retained, while the others were removed to avoid double-counting.

In addition, no inconsistencies were detected in any of the temporal or structural constraint checks described in Chapter 3. This indicates that the logs were internally coherent and suitable for subsequent analyses.

4.1.2.1 General Log Outlier Detection

To identify users with unusual activity patterns that could indicate logging errors, automated activity, or atypical usage behavior, a general outlier screening procedure was conducted across the combined app and portal logs.

Four summary indicators were computed for each participant: the total number of recorded events, the average number of events per active day, the median time gap between consecutive events, and the 95th percentile of inter-event gaps. These metrics were selected to capture both the overall volume of activity and the temporal characteristics of platform use.

Potential outliers were identified independently for each metric using the $3\times$Interquartile Range (IQR) rule. This more conservative threshold was chosen to focus on clearly extreme observations while reducing the likelihood of flagging genuinely highly engaged users.

The total number of events and the median inter-event gap each identified two users as outliers, while the average number of events per active day and the 95th percentile of inter-event gaps flagged four and five users, respectively. Across all indicators, 13 unique users were identified by at least one criterion. Figure 4.1 presents a scatter plot of the relationship between the total recorded events for both the application and the portal and median gap between each event for each participant. The identified outliers are highlighted in red.

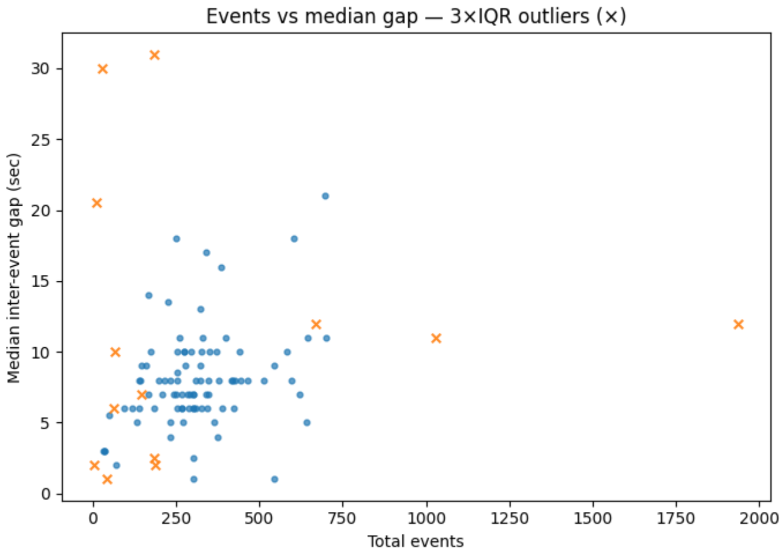


Figure 4.1: Relationship between total recorded events and median inter-event gap for each participant. Users flagged as outliers are highlighted.

Each identified case was manually inspected to determine whether the observed behavior reflected data quality issues or implausible activity patterns. No evidence of logging errors, duplicated activity, or non-human behavior was identified. Consequently, all users identified as outliers in this step were retained in the dataset for subsequent analysis.

4.2 Interaction Data Analysis

Following the procedures detailed in Section 3.5, descriptive results on user engagement with the app and portal functionalities and the ways in which these were navigated are presented in this section.

4.2.1 App Notifications

Notifications serve as a key engagement mechanism within the app, intended to encourage users to return to the platform and complete rating activities. Their effectiveness was examined across multiple dimensions, including delivery volume, response behavior, temporal patterns, conversion to app usage and rating activity, and potential indicators of notification fatigue.

Volume and response rates. Among all logged operations, the largest number were notification opens (4,117), while only two occurrences of deletes were noted. A significant quantity of events (2,106) were categorized as unknown, indicating cases where the user made no response to the notifications, likely showing disinterest.

Upon examination by the source (Table 4.3), clear differences in user behavior emerge. Notifications from the device achieved an open rate of 100%, with no deletions or unknown operations.

In comparison, notifications sent from the app were opened only around 35% of the times, with over 65% marked as unknown. Deletion incidents stayed negligible in both categories.

These results suggest that the source of a notification strongly influenced engagement. Table 4.3, which visualizes the breakdown by source, highlights this contrast, showing that device-sourced notifications consistently drove user interaction, while app-sourced ones often failed to.

Table 4.3: Operation breakdown by source.

Source	Open Rate	Delete Rate	Unknown Rate
app	0.346	0.00062	0.653
device	1.000	0.00000	0.000

Temporal patterns. In order to check when users exhibited greater responsiveness, notification openings were compiled based on the day of the week and hour. The heatmap in Figure 4.2 indicates clear temporal spikes: the peaks of opens were regularly noted at approximately 9:00 and 18:00 daily. Additional peaks were observed between 13:00 and 15:00, as well as near 19:00. There were slight differences between weekdays, although Wednesdays had a slight higher count of opens in relation to the other days. These patterns suggest that timing notifications for morning and evening peaks could improve engagement.

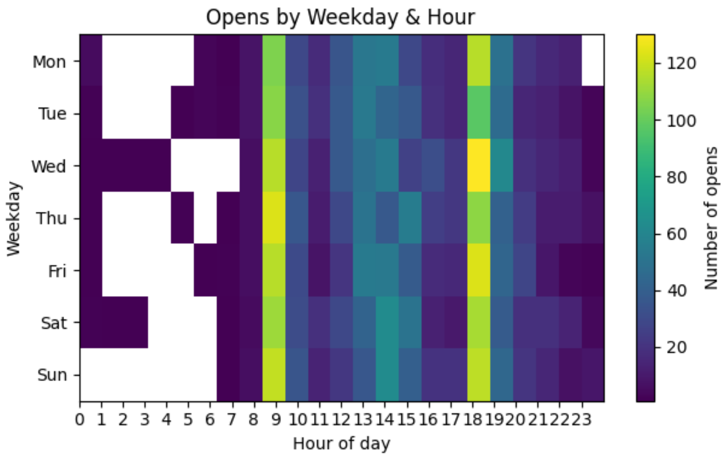


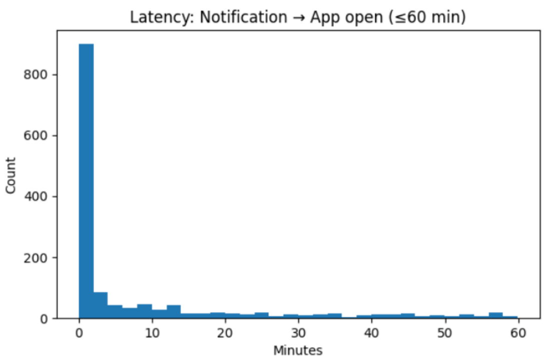
Figure 4.2: Heatmap of notification opens by weekday and hour.

Notification to app engagement. To assess whether notifications were related to subsequent app usage, the proportion of notifications followed by an *open* or *resume* event within the next 60 minutes was measured. The conversion rate was 0.235, indicating that nearly one in four notifications led to app usage.

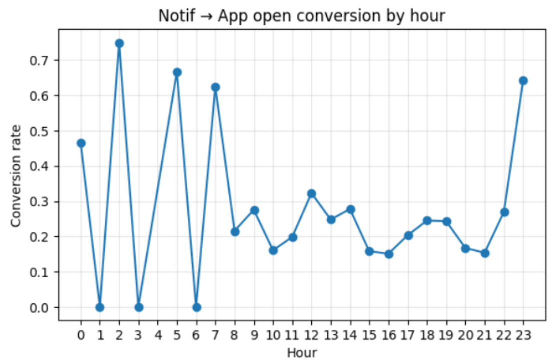
Figure 4.3a displays the latency distribution for notification to app opens. Most conversions happened immediately, with the highest density occurring in the first minute. The rate subsequently declined significantly, featuring only a few events within 1–10 minutes and just minimal activity up to the 60-minute limit.

Temporal patterns in conversion are shown in Figures 4.3b and 4.3c. By hour of day, conversion was highly volatile during nighttime hours (23:00–07:00), largely reflecting the small number of notifications opened at those times. This can be seen clearly in 4.3b, with similar values and without extreme peaks between 08:00 and 22:00, the period when there are more notification opens. Since there aren't almost any opens outside this period, the graph shows various peaks such as the ones at 01:00, 05:00 and 07:00. Between 08:00 and 22:00, a midday peak appeared around 12:00, followed by a smaller dip at 16:00. On weekdays, conversion rates were fairly consistent (0.23–0.25) from Monday through Saturday, with Tuesday and Friday showing slightly higher conversion rate. Sunday stood out with the lowest conversion rate (0.20), suggesting weaker re-engagement on that day.

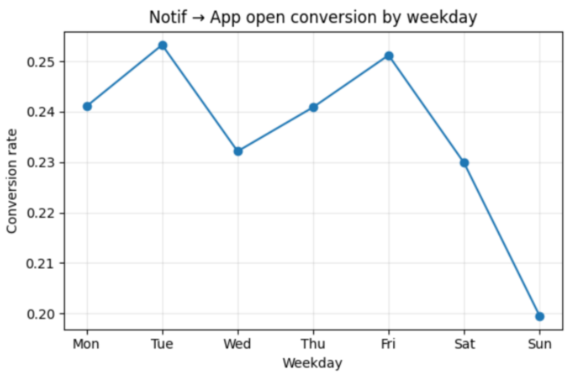
These results confirm that notifications act as effective short-term reactivation prompts, with most responses occurring within minutes and some variation across time of day and day of week.



(a) Latency distribution of notification → app-open conversions (< 60 min).



(b) Notification → app-open conversion rate by hour of day.



(c) Notification → app-open conversion rate by weekday.

Figure 4.3: Notification → app-open conversions: latency (a), hourly patterns (b), and weekday patterns (c).

Notification to Rating conversion. Beyond app re-engagement, we also examined whether notifications triggered subsequent rating events. Overall, the conversion rate within 60 minutes was very high at 0.921, indicating that nearly all rating events were preceded by a notification.

The latency distribution (Figure 4.4) shows that these conversions were immediate: virtually all rating events occurred within the first minute after a notification, with no meaningful activity observed beyond that point. This pattern suggests that when ratings did follow notifications, the response was instantaneous rather than delayed, highlighting a strong but narrow link between the two behaviors.

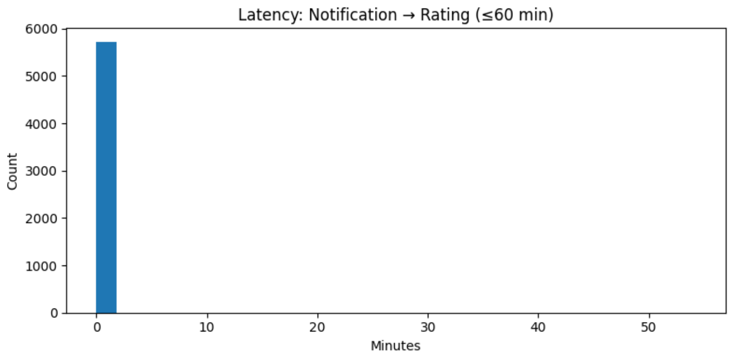


Figure 4.4: Latency distribution of notification $\rightarrow$ rating conversions ($<$ 60 min).

Engagement fatigue. The connection between the amount of notifications received and open rates was analyzed at a user level (Figure 4.5). Open rates for users who received between 0 and 100 notifications stayed fairly consistent, ranging from 0.6 to 0.8. Beyond this range, the lowess smoothing curve suggests greater variability: open rates declined sharply toward 0.3 near 200 notifications, rebounded to around 0.7 at 300, and then dropped again to about 0.3 at 500 notifications.

Although these changes suggest possible fatigue impacts at increased volumes, it is crucial to emphasize that the count of users getting over 100 notifications was very low. Consequently, the apparent decreases past this limit should be interpreted carefully. Overall, the evidence suggests stable engagement under typical volumes, with possible, but inconclusive signs of diminished responsiveness at extreme levels.

Summary. Overall, the notification analyses suggest that notifications are an effective mechanism for driving engagement with the mobile application, particularly for prompting rating activity. Device-delivered notifications consistently outperformed app-generated notifications, and most successful responses occurred immediately after delivery. While user responsiveness varied throughout the day, there was little evidence that the notification volumes observed in this study produced substantial notification fatigue. Together, these findings indicate that notifications played an important role in supporting user interaction with the platform.

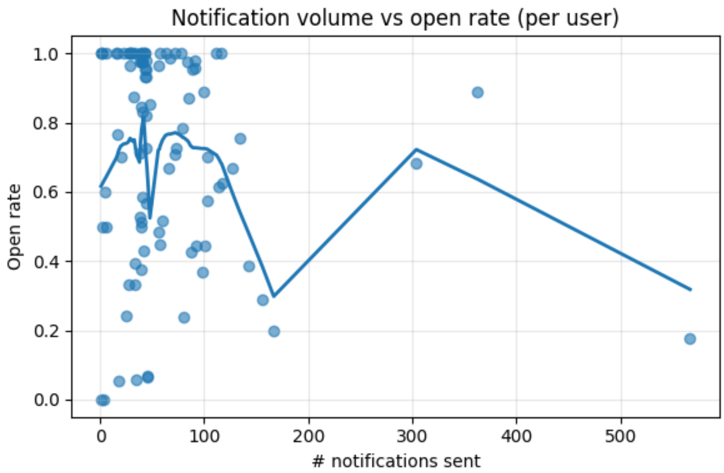


Figure 4.5: Scatter plot of notification open rate vs. notifications received (user-level), with lowess smoothing.

4.2.2 App Ratings

Rating behavior was analyzed in terms of its total volume and distribution across different sources. The findings are shown below.

Volume and source rates. The analysis of rating events uncovered a total of 4,941 ratings. As depicted in Table 4.4, most of these occurred following a notification from the device (60%). Ratings occurring after a notification from the app made up only 22%, whereas manual ratings comprised the least portion at 18%.

Table 4.4: Distribution of ratings by source.

Source	**Rate (%)**
device	60
app	22
manual	18

Summary. Overall, the rating analysis suggests that user rating activity was strongly shaped by the notification system. Most ratings were submitted following notification prompts, while manually submitted ratings were less frequent, indicating that spontaneous engagement with the rating feature was comparatively uncommon. Taken together, these findings highlight once again the central role of notifications in sustaining interaction with the application's self-monitoring functionality.

4.2.3 App Records

Data regarding the records, the feature that allowed users to write information about their food intake, sleep, and ongoing thoughts, were analyzed to characterize their volume, textual properties, and associations with sleep. The aim was to better understand how users engaged with self-reporting features of the app, and whether these records related to other outcomes such as mood ratings. Three main analyses were conducted: the volume of records by type, the characteristics of text-bearing records, and the distribution of sleep durations.

Volume by type. The analysis of records revealed a total of 101 entries, distributed unevenly across record types. As shown in Table 4.5, Food was the most frequent type with approximately 48% of the records being of this type, followed by Sleep with 40%. Thought records were comparatively rare, with only 12% of the records of this type.

This distribution highlights clear differences in how users engaged with the record functionality. Food and Sleep records accounted for the majority of submissions, whereas Thought records were comparatively uncommon. This may suggest that users were more willing to engage with concrete and routine self-monitoring tasks than with more reflective activities requiring greater cognitive effort.

Table 4.5: Distribution of records by type.

Source	**Rate (%)**
Food	48
Sleep	40
Thought	12

Text characteristics. For record types containing text (Food and Thought), the character length distribution of records was analyzed separately (Figures 4.6 and 4.7).

In Food records, most entries were short, with a higher peak between 7 and 10 characters, indicating that users generally recorded brief, practical notes.

In contrast, Thought records demonstrated substantially greater variability in length, ranging from very brief entries to considerably longer reflections. This may indicate that the effort and level of reflection involved in thought recording differed substantially across users.

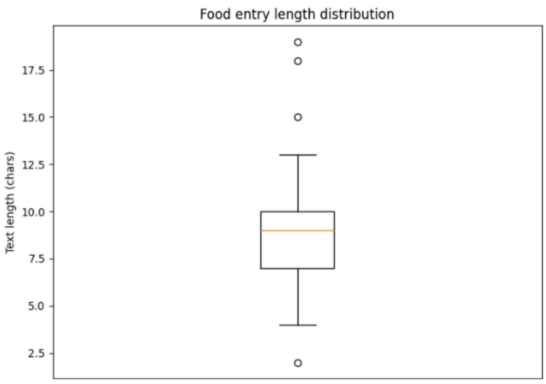


Figure 4.6: Distribution of text lengths for Food records.

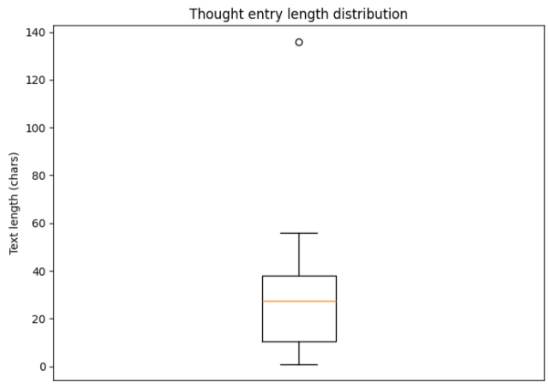


Figure 4.7: Distribution of text lengths for Thought records.

Sleep durations. Sleep records were analyzed separately because of their quantitative nature. Figure 4.8 displays the distribution of durations (limited to 5–15 hours because of an outlier entry near 30 hours). Most records were grouped around 8 to 10 hours, featuring a secondary peak around 9 hours and another between 6 and 7 hours.

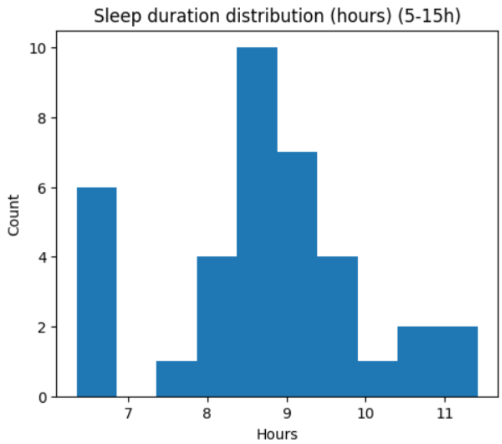


Figure 4.8: Distribution of sleep durations (minutes).

Summary. The record analysis revealed relatively limited use of the self-reporting functionality compared with other app features. Users engaged more frequently with concrete and routine record types, such as Food and Sleep, than with the more reflective Thought records. The textual characteristics of records further suggest differences in the level of effort required across record types, with Thought records showing substantially greater variability in length. Overall, these findings provide insight into how participants used the app's self-monitoring features and may help inform future improvements to their design and integration within the intervention.

4.2.4 App Navigation

User navigation within the app and portal was analyzed to characterize session structure, page usage, and flow patterns. Sessions were defined as sequences of activity separated by more than 30 minutes of inactivity, with very short sessions under 5 seconds excluded as accidental opens. The analyses below explore navigation at multiple levels of granularity, from page dwell times to transition patterns and higher-order sequence mining.

Page level analysis. As illustrated in Figure 4.9, the Rate page significantly outperformed all other pages regarding visits and dwell time. Coming next are the Notifications, Messages, and Message-Detail pages, whereas ones like Config, Records, and About received a comparatively low number of visits.

Median dwell times shown in Figure 4.10 offered another perspective. Even though Rate recorded the longest dwell time, the differences relative to other pages were slight. Notifications and Messages saw many visits but short dwell times, suggesting more temporary usage.

Together, these findings indicate that navigation was concentrated around the rating functionality, which accounted for most interactions, while lower-frequency sections tended to involve comparatively longer engagement once accessed. This suggests a distinction between frequently repeated routine actions and less frequent but more focused interactions.

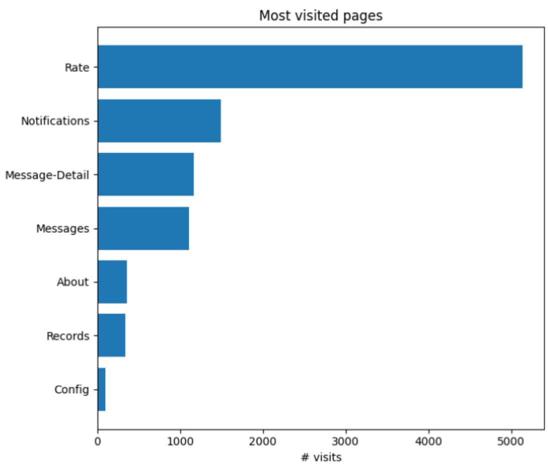


Figure 4.9: Most visited pages in the app.

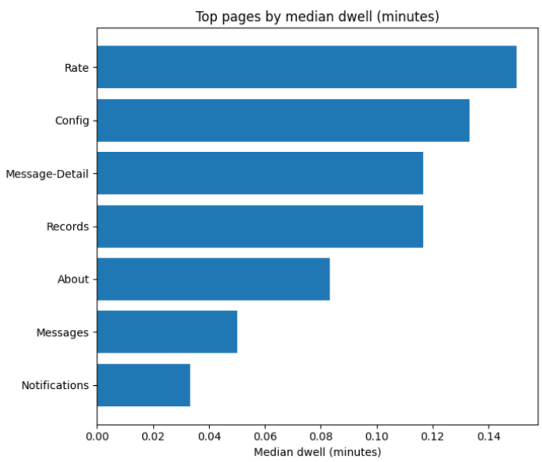


Figure 4.10: Median dwell time per page (minutes).

Session structure. Metrics at the session level were utilized to define broader patterns in app engagement. The session duration distribution (Figure 4.11) indicated that almost all sessions were shorter than one minute. A lesser proportion lasted 1–3 minutes, whereas longer sessions were infrequent and minimal in occurrence.

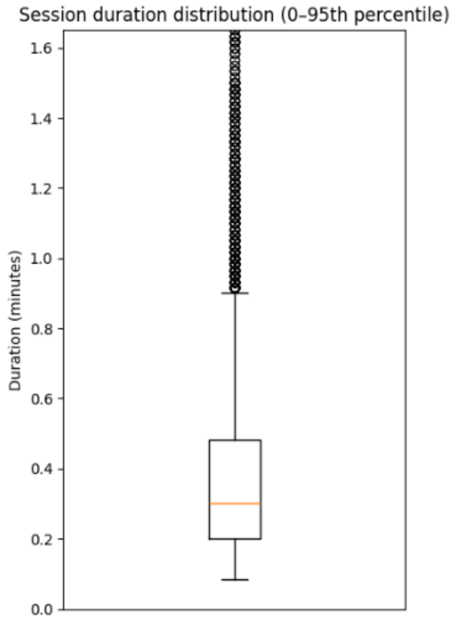


Figure 4.11: Distribution of session durations (minutes).

Page navigation within sessions exhibited a comparable trend of conciseness. As illustrated in Figure 4.12, nearly 3,000 sessions included just one page view, whereas an additional 1,300 had two pages. Sessions that included over two pages were relatively rare. The distribution of unique pages per session (Figure 4.13) echoed this result, with the vast majority of sessions restricted to one or two distinct screens.

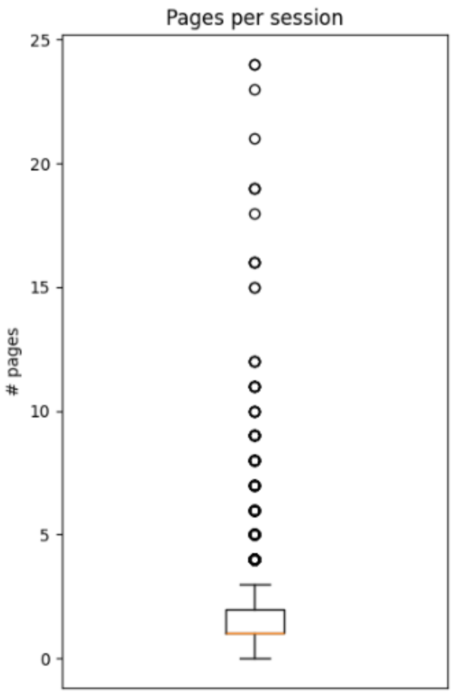


Figure 4.12: Distribution of pages per session.

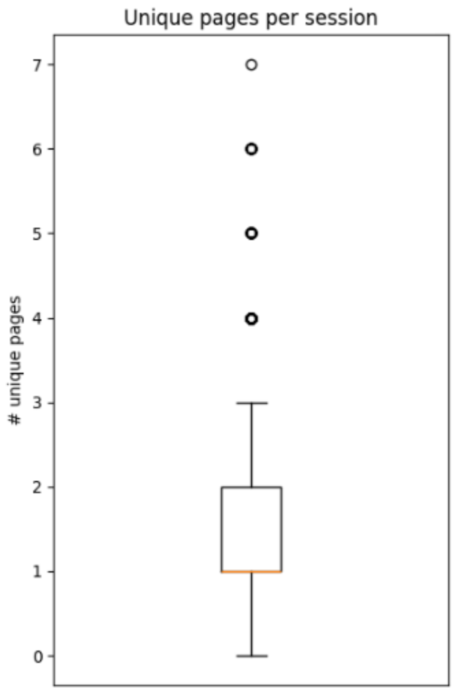


Figure 4.13: Distribution of unique pages per session.

Long-term patterns provided additional insights in addition to individual sessions. Each user's longest streak of consecutive active days is shown in Figure 4.14 with most users having streaks between 0 and 15 days. A smaller subset of users maintained streaks that lasted 20–30 days, while the distribution peaked at around 15 days. On the other hand, inactivity gaps (Figure 4.15) were typically short: the most common maximum gap was only one day, and occurrences decreased by half at two days and by four days. Long stretches of idleness were rare.

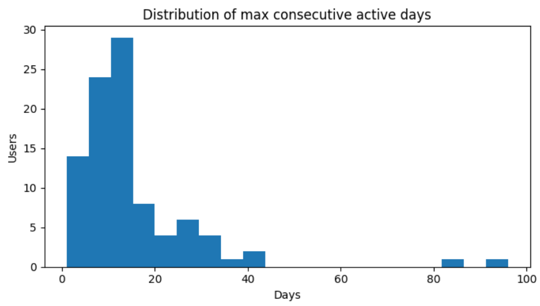


Figure 4.14: Distribution of maximum consecutive active days per user.

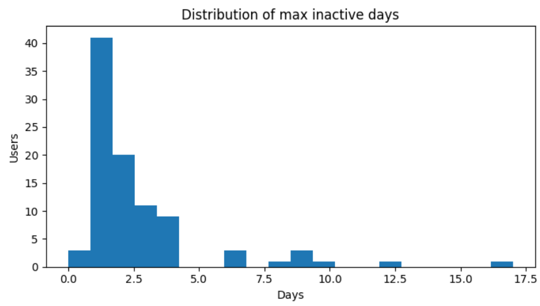


Figure 4.15: Distribution of maximum inactivity gaps (days) per user.

Overall, session-level metrics indicate that interactions with the application were generally brief and focused, with most sessions involving only one or two pages and lasting less than one minute. However, longitudinal indicators showed that many users returned regularly over extended periods, suggesting a pattern of frequent but lightweight engagement rather than prolonged single-session use.

Transitions and flows. Some complementary visual representations were used to assess the navigation dynamics. Page-level transition heatmaps (Figure 4.16) revealed a small number of prominent flows. Notifications → Rate (more than 1,000 occurrences) emerged as the dominant navigation flow. Together with the Sankey representation (Figures 4.17), this indicates that navigation was organized around a small number of recurring transitions, particularly those linked to notification-driven rating activity and message consumption.

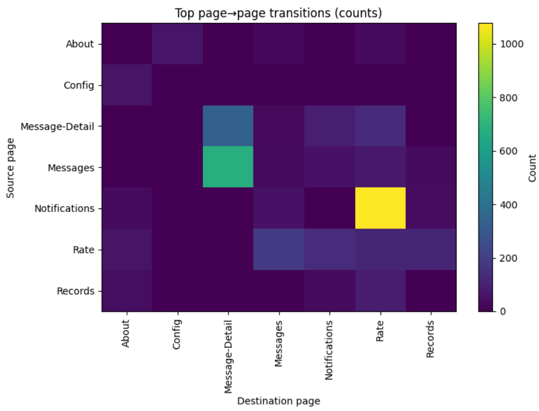


Figure 4.16: Heatmap of page-to-page transitions.

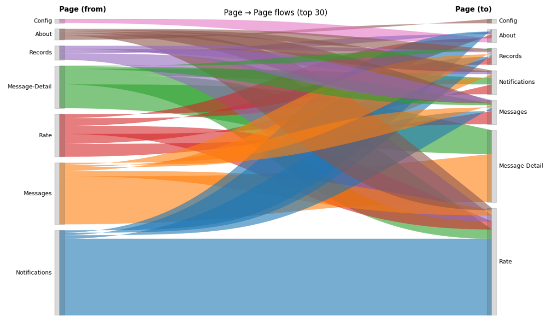


Figure 4.17: Sankey diagram of page-to-page transitions (top 30 flows).

Overall, the transition analysis suggests that user movement through the application followed a small set of highly recurrent pathways, with most interactions concentrated around ratings and message-related functionality.

Sequence mining. Finally, frequent navigation patterns were identified using sequence mining. N-grams of lengths $n \in \{2, 3\}$ were extracted using a sliding window. Table 4.6 shows examples of frequent sequences, highlighting recurring motifs.

Table 4.6: Most frequent page-level navigation sequences (top $n = 2$ and $n = 3$ n-grams).

Top $n = 2$		**Top $n = 3$**	
Pattern	**Count**	**Pattern**	**Count**
Notifications → Rate	1077	Message-Detail → Message-Detail → Message-Detail	220
Messages → Message-Detail	677	Rate → Messages → Message-Detail	113
Message-Detail → Message-Detail	334	Messages → Message-Detail → Message-Detail	112
Rate → Messages	184	Messages → Message-Detail → Rate	112
Rate → Notifications	138	Rate → Notifications → Rate	74
Message-Detail → Rate	130	Message-Detail → Notifications → Rate	72

At the page level, several key patterns emerged. The most frequent transition was notification-triggered rating: the bigram Notifications → Rate occurred 1077 times, far outnumbering the reverse (Rate → Notifications, 138). This imbalance supports the previous findings, stating that notifications serve as direct entry points to the rating process, with many sessions ending shortly afterward. Rate also functioned as a central hub, receiving significant traffic, particularly from Notifications, and redirecting users to other sections like Messages and Records.

The extracted sequences indicate that the most common navigation paths remained short and repetitive, rarely extending into longer interaction chains. This reinforces the broader pattern observed throughout the navigation analysis: app usage was predominantly composed of concise and goal-oriented interactions.

Summary. Overall, the navigation analysis suggests that interaction with the mobile application was highly structured around a small set of recurring behaviors. Most activity was concentrated on the rating functionality, with notifications acting as an important entry point into the application. Sessions were generally brief and involved limited exploration, yet many users returned consistently across multiple days. Transition analysis, Markov modelling, and sequence mining all pointed to the same conclusion: engagement was characterized more by frequent and routine interactions than by long or complex navigation paths. These findings provide insight into how users naturally engage with mobile mental health interventions and may help guide future improvements in interface design and feature integration.

4.2.5 Portal Logins

Patient logins in the web portal were analyzed to characterize patterns of access and engagement over time. In particular, login behavior was examined with respect to overall activity volume and its temporal distribution across days and hours.

Daily logins and intensity. The first analysis focused on login frequency throughout the observation period. Figure 4.18 shows daily login counts, which were generally distributed across the study window, although a noticeable decline in activity occurred between November 2021 and mid-December 2021. Outside this period, login activity displayed recurring peaks, indicating that portal use remained relatively consistent over time.

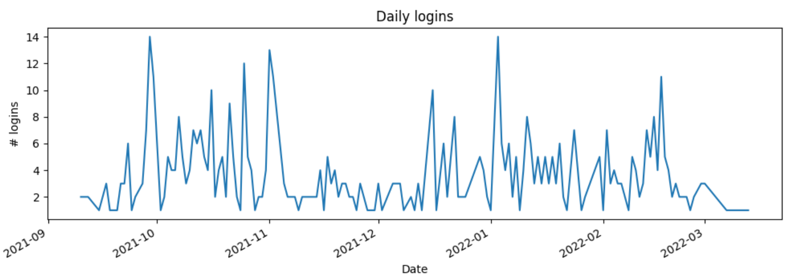


Figure 4.18: Daily portal login counts over the observation period.

Temporal patterns of access are summarized in 4.19 and 4.20. The heatmap of logins by weekday and hour reveals a clear concentration of usage during typical daytime and working hours, with activity peaking on Mondays around 13:00. As expected, login activity remained consistently low during night-time periods (00:00–07:00).

The weekday distribution indicates that Monday was the most active day, followed closely by Thursday. Tuesday and Wednesday showed slightly lower activity, while Friday presented a more noticeable decline. Weekend usage was lowest overall, with both Saturday and Sunday contributing relatively few logins.

The hourly distribution highlighted several recurring peaks across the day. The highest login frequency occurred around 10:00, followed by smaller peaks at approximately 13:00–14:00 and again near 19:00, suggesting that some users accessed the platform during lunch periods and evening hours.

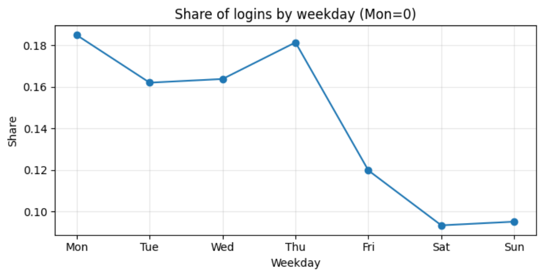


Figure 4.19: Share of portal logins across weekdays.

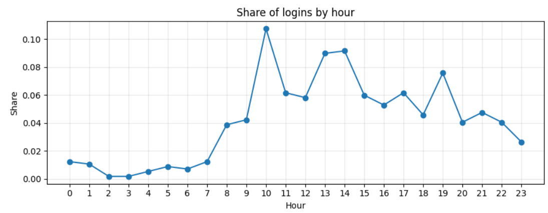


Figure 4.20: Share of portal logins by hour of day.

Summary. Overall, portal usage followed structured daily schedules rather than occurring uniformly throughout the day. Users predominantly interacted with the platform during daytime periods and weekdays, suggesting that portal engagement was often integrated into regular routines. The concentration of activity around working hours and the limited night-time usage indicate that the portal was primarily accessed in planned rather than spontaneous moments.

4.2.6 Portal Modules

Module progression was analyzed to understand how users advanced through the therapeutic content available in the portal. The analyses focused on completion rates, time required for completion, and the extent to which users engaged with the available materials.

Completion funnel. To evaluate progression through the portal modules, completion funnels were constructed for each module (Figure 4.21). The funnel captured the proportion of users who were assigned, who started, who progressed (defined as viewing at least one page and completing one activity), and who ultimately completed the module.

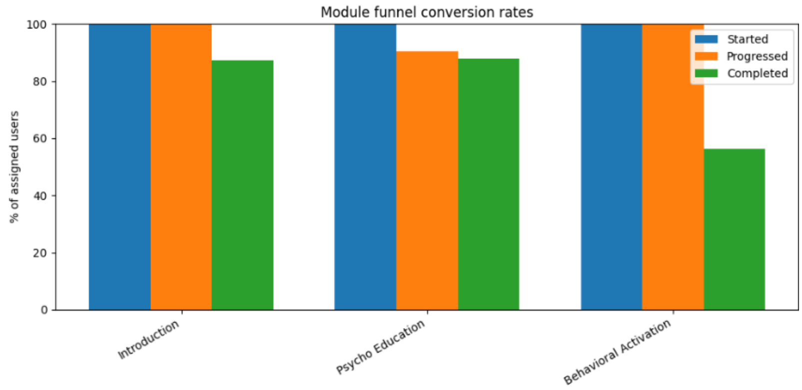


Figure 4.21: Completion funnels for selected portal modules.

Initial engagement was consistently high, with nearly all assigned users starting their modules. However, persistence differed across content stages.

For the *Introduction* module, progression remained stable across stages: 95 users started, all progressed, and 83 (87%) completed the module. The *Psycho Education* module followed a similar pattern, with 83 starts, 75 users progressing (90%), and 73 (88%) reaching completion. In contrast, the *Behavioral Activation* module exhibited substantially higher attrition. Although all 71 users who started progressed, only 40 (56%) ultimately completed the module.

These results suggest that sustaining engagement became more difficult as users progressed into more demanding therapeutic content.

Throughput time. The time required to complete each module was measured from the first recorded interaction. Figure 4.22 summarizes completion times through median and interquartile ranges.

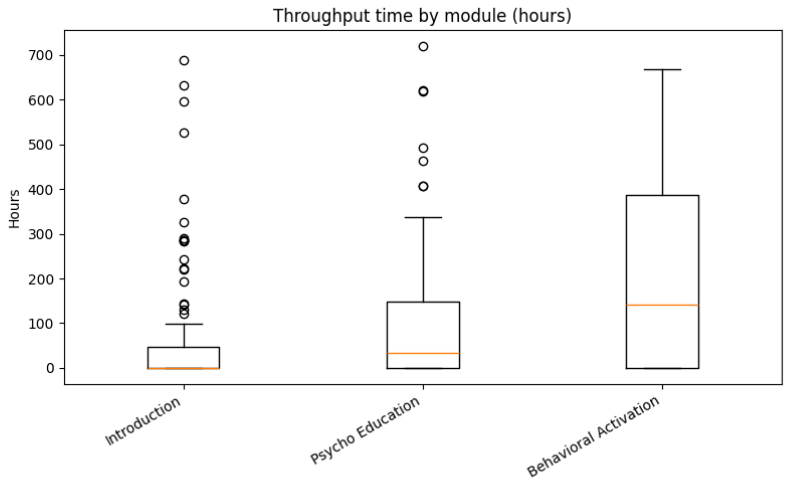


Figure 4.22: Throughput time (median and IQR) per module.

Completion times varied substantially across modules. The *Introduction* module was generally completed quickly, with a median completion time below one hour, although some users distributed their activity over nearly two days. Completion times increased in the *Psycho Education* module, with a median of approximately 32 hours and considerably greater variability.

The *Behavioral Activation* module required the longest time to complete, with a median duration close to six days and an upper quartile extending beyond sixteen days.

This progressive increase in completion time likely reflects both the growing complexity of the therapeutic content and a gradual reduction in user persistence throughout the intervention.

Module coverage. To assess how thoroughly users interacted with module content, the percentage of pages visited and exercises completed was calculated for each module (plots are provided in Appendix B.1).

Across all modules, page coverage remained consistently high, with most users viewing nearly all available content. Exercise completion, however, showed greater variability.

For the *Introduction* module, page coverage was nearly universal and exercise completion was generally high, although some users completed only part of the available activities. In the *Psycho Education* module, page coverage remained close to complete, but many users completed only around half of the exercises.

Among users who reached the *Behavioral Activation* module, exercise completion appeared stronger overall, with many users completing all activities. However, these patterns should be interpreted cautiously since this module was completed by a smaller subset of participants.

These findings suggest that users generally explored the available therapeutic content but were more selective in completing active exercises than in consuming informational material.

Summary. Overall, module progression followed a pattern of strong initial uptake followed by increasing attrition and longer completion times in later stages of the intervention. Users generally viewed most of the available content, but completing exercises required greater commitment and appeared more sensitive to disengagement. Together, these results indicate that maintaining active participation became progressively more challenging as the intervention advanced.

4.2.7 Portal Messages

Therapist messages were analyzed to assess whether they acted as triggers for subsequent portal engagement and progression through therapeutic content. Their potential influence was evaluated by examining whether message delivery was followed by login activity, interactions with module content, or progression-related events within predefined time windows. To distinguish message-driven behavior from users' normal activity patterns, post-message conversion rates were compared with baseline activity observed during equivalent periods preceding message delivery.

Message to login conversion. The first analysis examined whether receiving a therapist message increased the likelihood of logging into the portal within the following 24 hours. Only 3.5% of messages were followed by a login during this period, compared with a baseline login rate of

4.0% during the 24 hours preceding message delivery. This corresponds to a small negative lift of approximately -0.6 percentage points.

These results suggest that therapist messages did not substantially increase the probability of users returning to the portal in the short term.

Message to module activity conversion. A second analysis assessed whether messages were followed by interactions with therapeutic content. Specifically, we examined whether users visited any module page within 48 hours of receiving a message.

The observed conversion rate was 4.7%, compared with a baseline activity rate of 5.5% during the preceding 48 hours, resulting in an estimated lift of -0.9 percentage points. This indicates that module-related activity occurred slightly less frequently after receiving a message than during comparable baseline periods.

Message to progression events. To investigate whether therapist communications encouraged users to advance through the intervention, module starts and completions occurring within seven days of message delivery were examined.

Module starts were observed after 6.5% of messages, compared with a baseline rate of 9.9%, corresponding to a lift of -3.4 percentage points. Similarly, module completions occurred after 6.3% of messages, compared with a baseline rate of 8.4%, producing a lift of -2.1 percentage points.

Additional analyses of module-operation activity within 48 hours showed a conversion rate of 3.1%, compared with a baseline rate of 3.6%, yielding a lift of -0.5 percentage points.

Across all progression-related indicators, post-message activity was consistently lower than baseline levels, providing little evidence that therapist messages directly stimulated advancement through therapeutic content.

Summary. Overall, therapist messages showed limited effectiveness as direct triggers of portal engagement or module progression. Only a small proportion of messages were followed by subsequent logins, content interactions, or progression-related events, and in all cases the observed activity levels were slightly below users' baseline behavior. Unlike mobile notifications, which demonstrated a clear relationship with subsequent actions, therapist messages did not appear to function as strong standalone prompts for engagement. Their value may therefore lie more in providing support and maintaining therapeutic contact than in directly stimulating immediate platform activity.

4.2.8 Summary and Key Takeaways

The analysis of interaction data revealed distinct engagement patterns across the mobile application and the web portal, suggesting that both platforms fulfilled complementary roles within the intervention rather than functioning as interchangeable access channels.

For the mobile application, engagement was strongly driven by notifications and concentrated around a limited number of recurring actions, particularly mood rating activity. Device-originated notifications consistently achieved stronger user response than notifications generated directly by the application, indicating that notification delivery mechanisms substantially influence re-engagement effectiveness. Temporal analyses additionally showed that user responsiveness varied across the day, with higher activity observed during morning and evening periods. Together, these findings suggest that engagement with the application was largely externally prompted and organized around short interaction cycles.

The app navigation analyses reinforced this interpretation. Sessions were generally brief and focused, with most users entering the application to perform one or two specific actions before leaving. Rating-related flows dominated navigation behavior, particularly transitions from notifications directly into the rating process. At the same time, features such as records and more reflective content showed comparatively lower usage frequencies despite generating longer interaction times when accessed. This suggests that these features may support deeper engagement but currently play a secondary role in overall platform use.

In contrast, interaction with the web portal followed a more structured and content-oriented pattern. Login activity aligned closely with daily routines, concentrating during daytime hours and weekdays. Module analyses showed strong initial participation and high completion rates for early therapeutic content, but increasing attrition and longer completion times in later modules. Although users generally explored most of the available material, exercise completion became more selective as modules progressed, suggesting that maintaining active participation became increasingly challenging as the intervention demanded greater effort.

Therapist messages showed limited evidence of acting as direct engagement triggers. Message delivery was rarely followed by portal logins, module activity, or progression events at rates exceeding baseline behavior. Unlike mobile notifications, which demonstrated a clear and immediate relationship with subsequent user actions, therapist communications appeared to have a weaker and less direct influence on engagement patterns.

Taken together, these findings point to two complementary forms of engagement. The mobile application primarily supported frequent, lightweight, and externally prompted interactions that helped sustain regular contact with the intervention, while the web portal supported more deliberate and structured participation with therapeutic content. This distinction suggests opportunities for future improvements that strengthen the relationship between both platforms. In particular, the responsiveness and recurring usage patterns observed in the mobile application could be leveraged to better support engagement with the more demanding activities delivered through the portal, for example through stronger cross-platform integration, or mechanisms that encourage continued progression through therapeutic modules.

4.3 Correlation Analysis

This section examines the relationships between engagement metrics and intervention outcomes. Correlation analyses were conducted to explore whether engagement extracted from platform interactions were associated with changes in motivation and depressive symptoms over the course of the intervention. Additionally, associations between the changes in motivation and depression, and the other self-reported assessment measures, were also evaluated. In the heatmaps presented in this section, each cell displays the Spearman correlation coefficient (ρ), with an asterisk (*) marking associations that reached statistical significance ($p < 0.05$).

4.3.1 Self-reported Assessment Inter-Correlations

Before analyzing engagement metrics, correlations among the self-reported assessment measures were explored. Figure 4.23 presents the Spearman correlation matrix between usability and satisfaction outcomes (SUS and CSQ-I), baseline depressive symptoms (PHQ-9), adherence indicators, and changes in motivation and depressive symptom severity (SMFL difference and HADS-D difference).

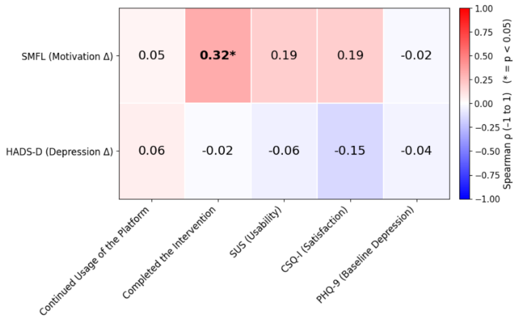


Figure 4.23: Spearman Correlations between SUS, CSQ-I, PHQ-9, adherence indicators and SMFL difference and HADS-D difference.

The strongest relationship was observed between intervention completion (adherence_2) and changes in motivation. Participants who completed the intervention generally demonstrated larger increases in motivation scores, suggesting that sustained participation is associated with more favorable motivational outcomes. This was also the only pair statistically significant according to the p-value.

Positive, although weaker, associations were also observed for usability (SUS) and satisfaction (CSQ-I). Participants reporting higher satisfaction and greater perceived usability tended to exhibit greater improvements in motivation, aligning with expectations that positive user experience may support continued engagement with treatment.

By contrast, adherence_1 (continued use of Moodbuster Lite throughout the study period) and baseline PHQ-9 scores showed minimal association with changes in motivation, indicating that remaining enrolled or entering the study with slightly different symptom levels was not strongly related to motivational change.

For depressive symptoms, represented by HADS-D difference scores, the strongest relationship emerged with satisfaction (CSQ-I). Participants who reported greater satisfaction tended to show larger reductions in depressive symptoms. Apart from this result, associations with the remaining measures were generally weak, suggesting limited direct relationships between baseline self-reported factors and symptom change.

4.3.2 Mobile Application Metrics

Figure 4.24 presents the Spearman correlations between mobile application engagement metrics and changes in motivation (SMFL difference) and depressive symptoms (HADS-D difference).

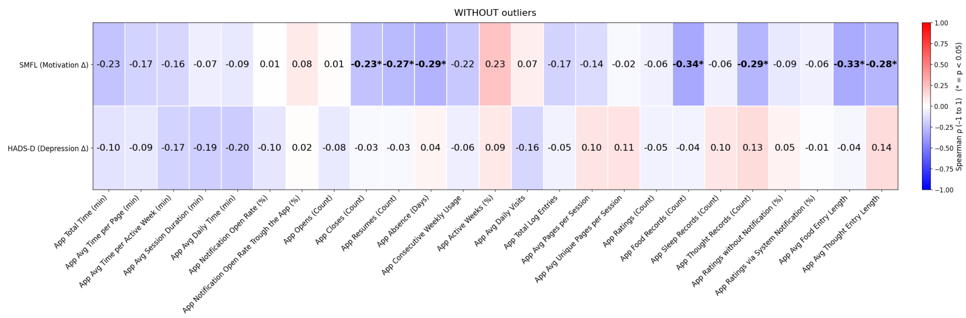


Figure 4.24: Correlation heatmap for the mobile application metrics and the difference on motivation and severity of depressive symptoms (SMFL difference and HADS-D difference)

Overall, relationships between app engagement and outcomes were modest, with most coefficients remaining within the interval $[-0.2, 0.2]$.

For the differences in motivation, many app metrics showed small negative associations with motivational improvement. The strongest effects appeared for metrics related to food and thought records, followed by some time-based and usage-based indicators. This pattern suggests that greater use of the app, specially the records feature, was not necessarily linked to increased treatment motivation. These associations, regarding the records, app opens, app closes and absence days were the only that showed to be statistically significant.

For the difference in depressive symptoms, time-oriented engagement measures displayed the most consistent negative associations. Average session duration, average daily time, average time spent in active weeks, and average daily visits all pointed in a similar direction, indicating that users who spent more time and interacted more frequently with the application tended to show slightly greater reductions in depressive symptoms.

Metrics related to food and thought records showed the opposite tendency, displaying positive correlations with the difference in depressive symptoms and suggesting weaker symptom improvement.

4.3.3 Web Portal Metrics

Figure 4.25 presents the correlations between web portal engagement metrics and and changes in motivation (SMFL difference) and depressive symptoms (HADS-D difference).

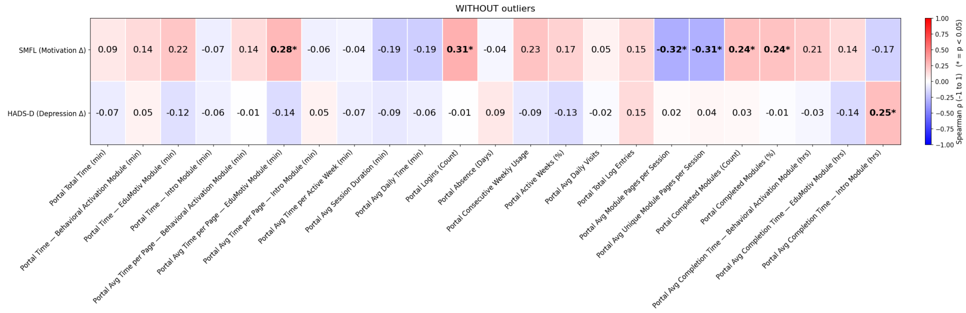


Figure 4.25: Correlation heatmap for the web portal metrics and the difference on motivation and severity of depressive symptoms (SMFL difference and HADS-D difference).

As with the mobile application, portal correlations were generally weak. However, the direction of associations with the motivation difference differed.

For the difference in motivation, most portal engagement measures showed positive relationships. Indicators of sustained participation, including login count, consecutive weekly usage, and total portal activity, were associated with greater improvements in motivation. Measures related to module progression, particularly completion time for the Behavioral Activation module, also showed positive associations.

Regarding the p-value analysis, for the difference in motivation, the average time per page in the EduMotiv module, the total logins and the number of completed modules are the ones statistically significant, pointing to an improvement in the motivation throughout the intervention. On the other hand, the average module pages and unique module pages per session were also statistically significant, but pointed to a reduction in the motivation.

Conversely, navigation-intensity metrics such as average pages per session, average unique pages per session, average session duration, and average daily time tended to correlate negatively with motivation improvement. These findings suggest that maintaining regular use over time may be more beneficial than concentrating engagement into fewer but longer sessions.

For the difference in depressive symptoms, the strongest associations again reflected consistency rather than intensity of engagement. Greater time spent during active weeks and higher percentages of active weeks following first login were associated with stronger symptom reduction, whereas more absence days showed the opposite pattern.

Additionally, longer completion times for the Introduction module were associated with weaker symptom improvement, suggesting that delays in progressing through early intervention content may be linked to poorer outcomes. This was also the only metric with an association statistically significant.

Taken together, the portal results emphasize the importance of sustained and timely participation throughout the intervention rather than isolated periods of intensive use.

4.3.4 Temporal Preference

To investigate whether engagement timing influenced intervention outcomes, users were grouped according to their predominant time of day of interaction. Temporal preference was defined as the time-of-day period with the highest login count, encoded as 1 (morning, 06:00–11:59), 2 (afternoon, 12:00–17:59), 3 (evening, 18:00–21:59), or 4 (night, otherwise).

Figures 4.26 and 4.27 show the distributions of SMFL difference and HADS-D difference across temporal preference groups.

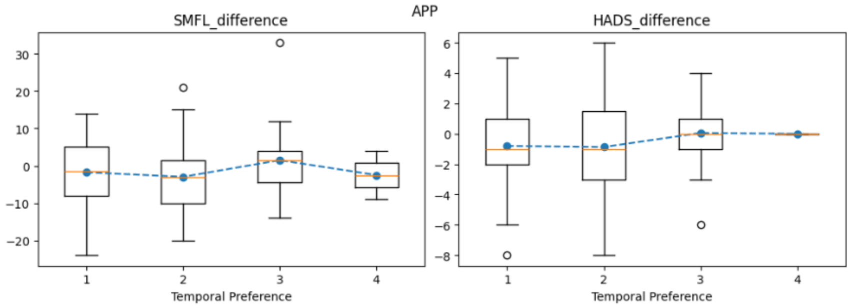


Figure 4.26: Distribution of SMFL_difference and HADS_difference across temporal preference groups for mobile application users.

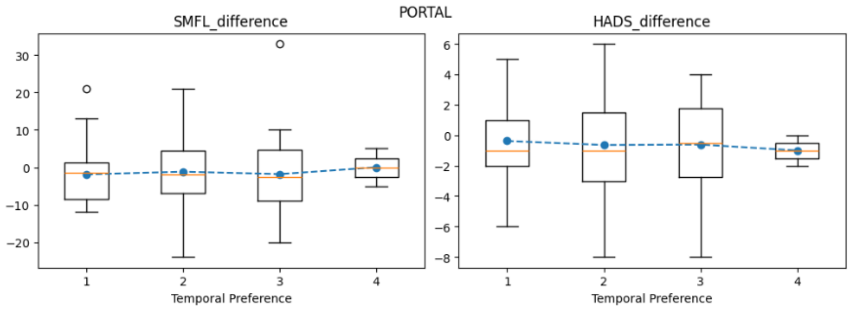


Figure 4.27: Distribution of SMFL_difference and HADS_difference across temporal preference groups for web portal users.

Differences between groups were small relative to within-group variability, and substantial overlap was observed across all distributions. No consistent evidence emerged suggesting that engaging predominantly during a specific time period was associated with better intervention outcomes.

These findings indicate that the timing of interaction itself was considerably less important than the overall consistency and pattern of engagement.

4.3.5 Summary and Key Takeaways

Overall, the correlation analyses revealed that relationships between engagement metrics and intervention outcomes were generally weak. No individual metric showed a strong association with either motivation or depressive symptom change. Nevertheless, several recurring patterns emerged.

Although a small number of correlations reached conventional significance thresholds, these results should be interpreted cautiously given the large number of comparisons performed and are best considered alongside the magnitude and direction of the observed associations.

Within the mobile application, engagement with reflective features and certain usage measures tended to relate negatively to motivational outcomes, whereas time spent and regular interaction were modestly associated with greater reductions in depressive symptoms.

For the web portal, sustained participation appeared more important than intensive use. Users who logged in consistently, progressed through intervention content, and maintained engagement over time tended to exhibit greater motivational gains and better symptom outcomes.

Self-reported assessments provided additional context. Improvements in motivation were most strongly associated with completing the intervention, while satisfaction and usability were positively related to motivational gains. For depressive symptoms, participant satisfaction emerged as the clearest associated factor.

Temporal preference showed minimal influence, reinforcing the idea that consistency and continuity of engagement were more important than the specific time of day users interacted with the intervention.

Overall, the findings support viewing engagement as a multidimensional construct. Although individual associations were generally modest and only a limited number reached statistical significance, the results consistently suggest that sustained and regular participation throughout the intervention may be more relevant for therapeutic outcomes than isolated periods of intense activity.

4.4 Predictive Modelling

In addition to the correlation analysis, predictive modelling was performed to evaluate whether engagement metrics extracted from the mobile application and web portal could be used to identify participants who experienced improvement in depressive symptoms during the intervention.

Separate logistic regression models were trained for the mobile application and web portal datasets. The prediction target was based on the binarised HADS-D difference score, distinguishing participants who showed improvement from those who did not. To reduce class imbalance effects, balanced subsets were used for model training. The mobile application dataset included 71 participants (35 improved, 36 non-improved), while the web portal dataset included 70 participants (37 improved, 33 non-improved). These numbers differ across platforms because not all participants used both components of the intervention; although access was provided to both the mobile application and the web portal, engagement varied naturally, with some users only interacting with one of the two platforms.

4.4.1 Mobile Application

The predictive model was estimated from the mobile application engagement metrics showed limited predictive ability. Performance remained close to chance level, achieving an accuracy of

0.494 ± 0.095, an F1 score of 0.489 ± 0.087, and an AUC of 0.482 ± 0.114.

These results indicate that engagement measures extracted from the mobile application alone were insufficient to reliably distinguish users who improved from those who did not. The relatively large variability across validation folds further suggests instability in model behavior across different subsets of participants.

Although the model did not achieve meaningful predictive performance, the estimated coefficients (Figure 4.28) still provide exploratory insight into the types of engagement that appeared more frequently among users classified as improved.

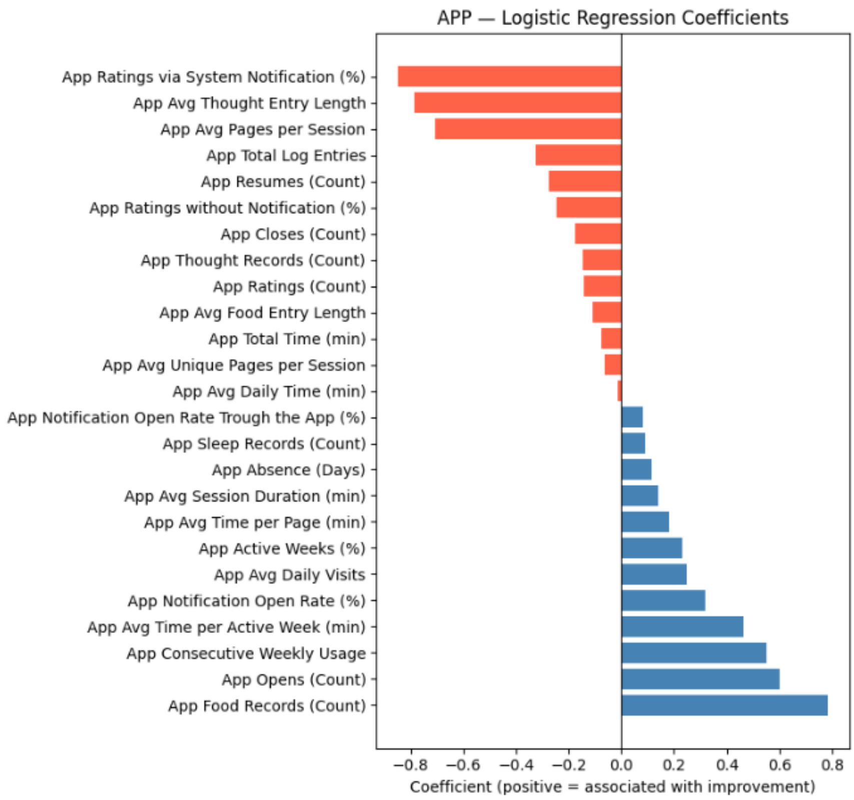


Figure 4.28: Logistic regression coefficients for mobile application engagement metrics.

The strongest positive coefficients were associated with food record count (0.78), app open count (0.60), consecutive weekly usage (0.55), and average time spent during active weeks (0.47). Smaller positive effects were also observed for notification open rate and average daily visits.

Taken together, these variables suggest that users who returned to the application more frequently and maintained engagement across multiple weeks were somewhat more likely to be classified into the improved group.

Negative coefficients showed a different pattern. The largest negative values were associated with the percentage of ratings triggered by system notifications (−0.85), average thought record length (−0.79), average pages per session (−0.71), total log entries (−0.32), and resume count (−0.28).

These patterns may indicate that more reactive behaviors driven primarily by external prompts, as well as longer and potentially more fragmented sessions, appeared more frequently among users who did not show improvement.

However, because overall predictive performance remained weak, these coefficient patterns should be interpreted carefully.

4.4.2 Web Portal

The predictive model based on web portal engagement metrics produced similarly limited results. Performance metrics remained close to random classification, with an accuracy of 0.514 ± 0.053, an F1 score of 0.569 ± 0.106, and an AUC of 0.457 ± 0.115.

These results suggest that portal interaction behavior alone did not contain sufficient information to reliably predict changes in depressive symptoms.

Despite the limited predictive performance, examining model coefficients (Figure 4.29) provides additional context regarding the engagement characteristics most frequently associated with improvement.

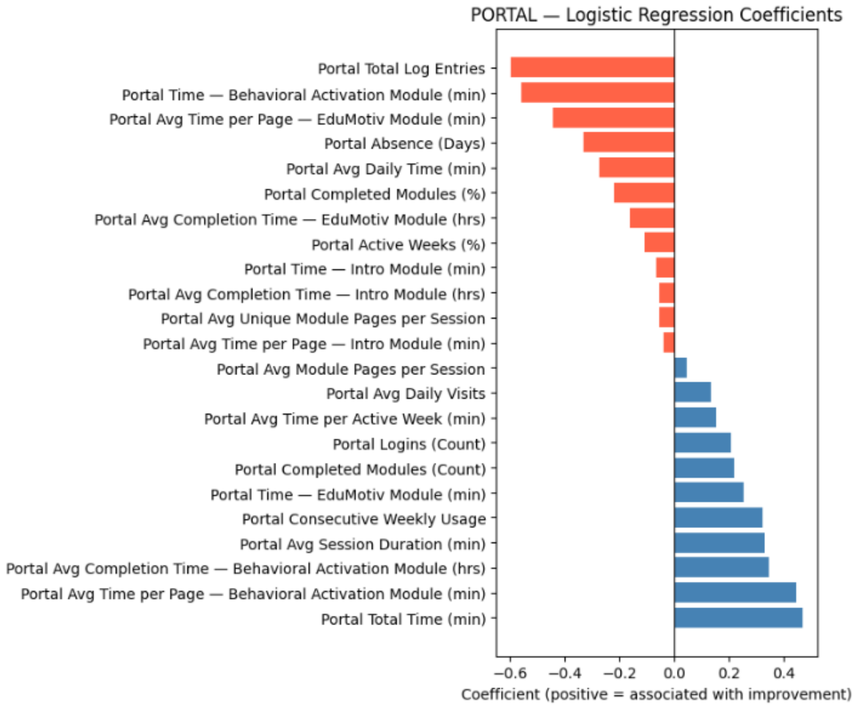


Figure 4.29: Logistic regression coefficients for web portal engagement metrics.

The strongest positive coefficients were observed for total time spent on the portal (0.47), average time spent on Behavioral Activation module pages (0.45), average completion time for the Behavioral Activation module (0.35), average session duration (0.33), and consecutive weekly usage (0.32). Additional positive contributions were observed for completed modules and login frequency.

These results broadly align with the earlier correlation analysis and suggest that users who sustained engagement with therapeutic content over time were somewhat more likely to belong to the improved group.

The strongest negative coefficients were associated with total log entries (-0.60), total time spent in the Behavioral Activation module (-0.56), average time spent on Psycho Education pages (-0.44), absence days (-0.33), and average daily time (-0.28).

This pattern again suggests that consistency of participation may be more informative than simply accumulating large amounts of activity or spending plenty of time in isolated periods.

As with the mobile application model, these coefficient patterns should be interpreted cautiously due to the limited predictive performance.

4.4.3 Overall Assessment

Overall, neither predictive model achieved meaningful classification performance. Across both platforms, accuracy and AUC values remained close to what would be expected under random guessing.

Several factors may explain these findings. First, the available sample size was relatively small for predictive modelling and may not have provided sufficient statistical power to capture stable behavioral patterns. Second, therapeutic response is likely influenced by many factors beyond observable engagement behavior, including individual differences in treatment responsiveness, external circumstances, and psychological characteristics not captured in interaction logs. Finally, it could be hard to represent engagement engagement adequately through the extracted metrics alone.

Even though the models did not perform well as predictive tools, the coefficient patterns remained broadly consistent with findings from the correlation analysis. Across both the mobile application and web portal, indicators of sustained and repeated engagement over time appeared more frequently among positive coefficients, whereas simply generating large volumes of activity did not consistently correspond to improved outcomes.

Taken together, these findings suggest that the quality and regularity of engagement may be more informative than overall usage intensity. While interaction metrics alone were insufficient for reliable prediction, they still provided useful insights.

Chapter 5

Discussion

The results presented in the previous chapter revealed several distinct patterns of engagement across the Moodbuster mobile application and web portal. While users interacted regularly with both platforms, engagement was expressed in different ways depending on the platform and functionality involved. Furthermore, although several engagement measures showed associations with intervention outcomes, these relationships were generally modest and often depended on the specific type of engagement being considered. This chapter interprets these findings in relation to the five research questions, discusses their implications for engagement measurement and platform design, and considers the limitations of this work.

5.1 RQ1: How do users interact with a digital mental health platform, as reflected in access log data?

The interaction analyses revealed that the mobile application and the web portal played complementary but clearly differentiated roles throughout the intervention.

For the mobile application, user behavior was strongly influenced by notifications. Device-level notifications achieved considerably higher response rates than notifications originating directly from the application, and a large proportion of mood ratings occurred shortly after receiving notifications. Combined with the navigation findings, this suggests that many users interacted with the app reactively rather than initiating usage independently.

Sessions were generally brief, commonly lasting under one minute and involving only one or two page views. Rather than exploring different areas of the application, users appeared to enter with a clear objective and exit shortly afterward. Nevertheless, many participants maintained repeated usage across days and weeks, indicating that engagement was lightweight but relatively stable over time.

The web portal showed a substantially different usage profile. User activity was primarily centred around progression through therapeutic modules and completion of intervention content. Completion remained high during the earlier stages of the intervention but decreased noticeably for later modules. This reduction in persistence likely reflects both increasing content complexity and the greater effort required to continue participation over time, suggesting that sustaining engagement becomes progressively more difficult as interventions demand more active involvement from users.

Therapist messages appeared to have a comparatively limited effect on subsequent engagement. Unlike notifications in the mobile application, messages did not consistently trigger additional portal logins or progression through module content.

Taken together, these findings suggest that the application and portal fulfilled complementary roles: the app facilitated repeated and lightweight contact with the intervention, whereas the portal supported deeper and more structured therapeutic engagement.

5.2 RQ2: What metrics can be defined to quantify user engagement?

One of the central contributions of this work was the development of a multidimensional set of engagement metrics derived entirely from interaction logs, without the need for additional questionnaires or data collection beyond what was already generated by the platform itself. The proposed metrics covered several complementary dimensions, including time-on-platform, interaction volume and frequency, content access and navigation, task and performance outcomes, and usage behavior over time.

These metrics were intentionally designed to reflect the distinct roles of the two platforms within the intervention, as described above. The mobile application mainly supported short and recurrent interactions, focused on mood ratings and lightweight self-monitoring activities. In contrast, the web portal was designed around progression through therapeutic content and more structured forms of engagement.

This distinction proved important throughout the analyses because engagement manifested differently depending on platform context. Metrics that were informative for one platform were not necessarily meaningful for the other, reinforcing the importance of evaluating engagement as a context-dependent phenomenon rather than through a single generic measure.

5.3 RQ3: How do direct engagement metrics from interaction data correlate with indirect measures of user engagement?

The correlation analyses explored three types of relationships within the data. The first examined associations among the self-reported assessment measures themselves, considering whether baseline characteristics such as perceived usability (SUS), user satisfaction (CSQ-I), and baseline depression severity (PHQ-9) were related to changes in treatment motivation (SMFL difference) and depressive symptom severity (HADS-D difference). The second and third examined whether direct engagement metrics derived from the mobile application and the web portal, respectively, were associated with changes in both outcome measures. Since changes in depressive symptom severity are more directly indicative of therapeutic success, the findings regarding the HADS-D difference are discussed in the context of the following research question; the present section focuses on the associations with motivational change.

Associations between self-reported measures and motivational change were generally weak. Neither perceived usability nor user satisfaction showed strong relationships with improvements in motivation, suggesting that how users felt about the platform did not directly translate into greater motivational benefit. Baseline depression severity also showed limited association with motivational change, indicating that initial symptom levels alone were not a reliable predictor of how motivation would evolve during the intervention.

Regarding direct engagement metrics, the mobile application and the web portal showed notably different patterns in relation to motivational outcomes. Within the mobile application, several engagement indicators were associated with reductions in motivation over time, particularly metrics related to reflective activities such as food and thought records. In contrast, portal engagement metrics tended to show positive associations with motivational improvement, especially measures linked to sustained participation and progression through therapeutic content, suggesting that deeper involvement with structured intervention content may support the growth of treatment motivation.

5.4 RQ4: What is the relationship between user engagement and therapy outcomes?

The analyses relating engagement to depressive symptom outcomes indicate that the relationship between platform usage and therapeutic benefit is more complex than assuming that higher engagement automatically leads to better results.

Regarding depressive symptoms, a consistent pattern emerged across both platforms. Measures reflecting sustained participation over time, such as active weeks, consecutive usage, and continued involvement throughout the intervention, showed more favorable associations with symptom improvement than metrics representing isolated bursts of intensive activity. Conversely, indicators of interruption and irregular participation tended to relate to poorer outcomes. These findings support the idea that therapeutic benefit may depend more on maintaining regular engagement over time than on accumulating large volumes of activity within short periods.

Across both platforms, however, correlation magnitudes were generally small. This suggests that interaction behavior represents only one component within a broader set of factors influencing therapeutic change, and that the relationship between usage patterns and clinical improvement cannot be fully captured through log-derived metrics alone.

The predictive modelling results further support this interpretation. Logistic regression models were trained to assess whether engagement metrics could distinguish participants who improved in depressive symptoms from those who did not. Although neither the mobile application nor the web portal model achieved meaningful predictive performance, with results close to random classification, the coefficient patterns broadly aligned with the correlation findings. Metrics associated with continuity of participation appeared more frequently among positive predictors, while high interaction volume alone showed weaker and less consistent relationships with symptomatic improvement.

Overall, the findings reinforce the importance of distinguishing between different forms of engagement. Simply increasing usage may not be sufficient to produce therapeutic benefit, and encouraging regular, sustained, and meaningful participation is likely to be of greater value than maximising raw activity levels.

5.5 RQ5: What design improvements can be made to digital mental health platforms to enhance user engagement and therapy effectiveness, based on insights from user interaction data?

The interaction analyses revealed several opportunities for improving the design of digital mental health platforms.

One of the clearest findings was the strong dependence of mobile application usage on notifications. User activity increased substantially following notifications, indicating that reminders play an important role in prompting engagement. However, relatively little spontaneous use was observed outside these periods, suggesting that engagement may be overly reliant on external prompting. Future interventions should therefore continue to use notifications as a key engagement mechanism while also encouraging more self-initiated interactions. Features such as personalised goals or progress tracking may help users engage with the application more independently over time.

The notification analyses also provided practical guidance for delivery strategies. Notifications sent directly from the mobile device achieved higher open rates than those delivered through the portal, suggesting that device-based notifications should be prioritised whenever possible. Furthermore, user activity showed clear peaks around specific periods of the day, particularly in the morning and early evening. Scheduling notifications closer to these naturally active periods may increase the likelihood of user interaction and improve overall engagement.

The analyses additionally highlighted a disconnect between engagement with the mobile application and the web portal. Although participants had access to both components, usage patterns differed considerably between platforms, and engagement metrics from one platform were not consistently reflected in the other. This suggests an opportunity to strengthen the integration between the mobile application and portal experience. For example, progress made in one platform could be more prominently reflected in the other, creating a more unified intervention experience and encouraging users to engage with both components.

Therapist messages showed limited impact on subsequent portal logins or module progression, suggesting that current messaging practices were not effectively leveraged as a re-engagement mechanism. This represents an underutilised opportunity, given that personalised human contact may be a stronger motivator for continued participation in the intervention. Improving this could involve structuring therapist messages to include a concrete and specific prompt tied to platform activity, for example referencing a particular module the user has not yet completed or acknowledging a pattern in their recent mood ratings. Timing messages to coincide with periods of declining activity, rather than at fixed intervals, could also increase their relevance and effectiveness as a re-engagement trigger.

Within the web portal, participation gradually declined as users progressed through the intervention modules. While many participants started the later modules, fewer progressed and fewer still completed them. This pattern suggests that maintaining engagement becomes increasingly challenging as the intervention progresses. Design strategies aimed at improving retention in later modules, such as breaking content into smaller sections, reinforcing progress, providing milestone rewards, or introducing additional support, may help sustain participation throughout the intervention.

More broadly, the consistent association found between sustained and regular participation and improved outcomes suggests a general design principle: platform features should prioritise supporting continuity of use over time rather than maximising the intensity or volume of interaction within any single session. Design choices that help users maintain a lightweight but regular relationship with the platform, such as gentle reminders after periods of absence, and features that reward consistency rather than single-session depth, are likely to be more beneficial for therapeutic outcomes than those aimed at increasing the amount of content consumed in one sitting.

Overall, the findings point towards five key design priorities: optimising notification strategies, encouraging more spontaneous mobile application use, strengthening integration between platform components, making better use of therapist communication as a re-engagement mechanism, and improving retention throughout later stages of the intervention.

5.6 Limitations

Several limitations should be considered when interpreting these findings.

First, the sample available for outcome analyses was relatively small. Although the interaction logs contained rich behavioral information, the number of participants was low. This likely reduced statistical power and may partly explain the weak predictive performance observed.

Second, engagement itself is difficult to interpret from interaction logs alone. Higher activity does not necessarily imply more meaningful participation. Long sessions, large numbers of page visits, or frequent interactions may indicate productive engagement for some users but confusion, inefficiency, or increased effort for others. Likewise, the observed associations should not be interpreted causally.

A further limitation concerns the reconstruction of sessions. Session boundaries were inferred using inactivity thresholds and approximations required by the available log structure. Although practical and consistent, alternative session definitions could produce somewhat different estimates of engagement.

Methodologically, the analyses mainly focused on monotonic and linear relationships through Spearman correlations and logistic regression, so more complex and potentially non-linear relationships between engagement and outcomes may therefore remain unexplored.

Another limitation concerns the definition of the outcome used in the predictive analyses. Participants were classified as improved whenever their post-treatment HADS-D score was lower than their baseline score. While this definition captures any reduction in depressive symptoms, it does not distinguish between changes with different magnitudes.

Furthermore, the predictive models were developed without explicitly accounting for potential confounding variables. Factors such as baseline symptom severity, demographic characteristics, or other participant-specific variables may influence both engagement behavior and treatment outcomes. Consequently, the observed predictive relationships should not be interpreted as reflecting the independent effect of engagement alone.

Finally, the study was conducted using data from a single digital intervention platform and a specific participant population. Consequently, the findings may not generalize to other digital mental health interventions with different platform designs, therapeutic approaches, or user populations.

Overall, despite these limitations, the study demonstrates that interaction logs can provide meaningful insights into engagement behavior and highlights the importance of considering engagement as a multidimensional and context-dependent construct in digital mental health interventions.

Chapter 6

Conclusion

This dissertation investigated how user engagement in a digital mental health intervention can be measured using interaction log data and whether these engagement patterns are associated with therapeutic outcomes. Using interaction data collected from the Moodbuster platform, a comprehensive set of engagement metrics was developed for both the mobile application and the web portal. These metrics captured multiple dimensions of engagement, including activity volume, usage frequency, temporal behavior, content interaction, navigation patterns, and progression through intervention components.

A first contribution of this work was the design and implementation of a multidimensional engagement measurement framework based entirely on collected interaction logs. By relying exclusively on data already generated by the platform, the proposed metrics provide a scalable approach to monitoring engagement without increasing participant burden.

The analysis of interaction data provided valuable insights into how participants used the intervention. The mobile application and web portal were found to support different but complementary forms of engagement. Mobile app usage was characterized by short, frequent interactions strongly influenced by notifications, particularly around mood rating activities. In contrast, portal usage was more structured and centred on progression through therapeutic modules and intervention content. These findings contribute to a better understanding of how users interact with different components of digital mental health interventions.

The analyses relating engagement to therapeutic outcomes revealed that the relationship between platform usage and clinical improvement is complex. Although several recurring patterns emerged, particularly regarding the importance of sustained and consistent participation, most associations between engagement metrics and outcome measures were relatively weak. No individual metric emerged as a strong predictor of either motivational improvement or reductions in depressive symptoms.

The predictive modelling results reinforced these observations. Logistic regression models trained on engagement metrics from both platforms achieved performance close to random classification. These findings suggest that interaction data alone are insufficient to reliably predict therapeutic outcomes and that symptom improvement is likely influenced by a broader combination of clinical, personal, and contextual factors that are not captured by platform usage behavior alone.

Despite these limitations, the study demonstrates the value of interaction log analysis for understanding user behavior in digital mental health interventions. The engagement metrics and behavioral insights generated throughout this work provide a useful foundation for future monitoring systems, engagement analytics frameworks, and intervention design improvements. More broadly, the results highlight that engagement should be viewed as a multidimensional construct, where consistency, persistence, and the nature of interactions may be more informative than simple measures of activity volume.

6.1 Future Work

Several opportunities exist to extend and build upon the work presented in this dissertation.

A primary direction for future research is the use of larger datasets involving a greater number of participants. Larger samples would improve statistical power, enable more robust validation of the proposed engagement metrics, and support the development of more sophisticated predictive models.

Future studies could also integrate interaction logs with additional sources of information. Demographic characteristics, therapist interactions, and contextual information may provide complementary perspectives that help explain therapeutic outcomes more effectively than interaction data alone.

From a methodological perspective, more advanced analytical approaches could be explored. While this dissertation focused on correlation analysis and logistic regression, future work may benefit from techniques capable of capturing complex and non-linear relationships.

Future research should also investigate alternative definitions of treatment improvement. Rather than classifying participants as improved based on any reduction in depressive symptom scores, future studies could consider clinically meaningful change thresholds or reliable change indices. Such outcome definitions would better reflect therapeutic benefit and may improve the clinical relevance of predictive models.

Future predictive models should also incorporate potential confounding variables, such as baseline symptom severity, demographic characteristics, or other participant-specific factors. Accounting for these variables would help unravel the relationship between participation and treatment outcomes and allow for more accurate and interpretable predictions.

The interaction analyses also highlighted several practical opportunities for improving the Moodbuster platform itself. Future iterations could investigate alternative notification strategies, adaptive reminder schedules, stronger integration between app features, and mechanisms designed to encourage more spontaneous interaction rather than relying predominantly on external prompts. Similarly, additional support mechanisms could be explored to reduce attrition in later portal modules and sustain long-term participation.

Finally, the engagement metrics developed in this dissertation could be applied and validated in other digital mental health interventions. Comparing engagement patterns across different platforms, therapeutic approaches, and populations would contribute to a broader understanding of engagement in digital health.

Overall, this work provides a foundation for future research on behavioral engagement in digital mental health interventions and highlights both the opportunities and challenges associated with using interaction log data to understand and support therapeutic processes.

References

[1] N. Ademi, S. Loshkovska, and S. Kalajdziski. “Prediction of Student Success Through Analysis of Moodle Logs: Case Study”. In: *ICT Innovations 2019. Big Data Processing and Mining*. Ed. by S. Gievska and G. Madjarov. Cham: Springer International Publishing, 2019, pp. 27–40. ISBN: 978-3-030-33110-8. DOI: 10.1007/978-3-030-33110-8_3.

[2] A. A. El Alami, H. Ezzikouri, and M. Erritali. “Data Mining Algorithms for Knowledge Extraction from Web Log Files”. In: *Advanced Intelligent Systems for Sustainable Development (AI2SD’2019)*. Ed. by M. Ezziyyani. Cham: Springer International Publishing, 2020, pp. 118–128. ISBN: 978-3-030-36653-7. DOI: 10.1007/978-3-030-36653-7_12.

[3] K. Amarti et al. “Feasibility of Digital Cognitive Behavioral Therapy for Depressed Older Adults With the Moodbuster Platform: Protocol for 2 Pilot Feasibility Studies”. In: *JMIR Res Protoc* 11.10 (2022), e41445. DOI: 10.2196/41445.

[4] K. Aulakh, R.K. Roul, and M. Kaushal. “E-learning enhancement through educational data mining with Covid-19 outbreak period in backdrop: A review”. In: *International Journal of Educational Development* 101 (2023). DOI: 10.1016/j.ijedudev.2023.102814.

[5] I.D. Babić. “Machine learning methods in predicting the student academic motivation”. In: *Croatian Operational Research Review* 8.2 (2017), pp. 443–461. DOI: 10.17535/crorr.2017.0028.

[6] Rosa María Baños, Roberto Herrero, and María Dolores Vara. “What is the Current and Future Status of Digital Mental Health Interventions?” In: *The Spanish Journal of Psychology* 25 (2022), e5. DOI: 10.1017/SJP.2022.2.

[7] Amit Baumel and Fred Muench. “Heuristic Evaluation of Ehealth Interventions: Establishing Standards That Relate to the Therapeutic Process Perspective”. In: *JMIR Mental Health* 3.1 (2016), e5. DOI: 10.2196/mental.4563. URL: https://doi.org/10.2196/mental.4563.

[8] Jiska Borghouts et al. “Barriers to and Facilitators of User Engagement With Digital Mental Health Interventions: Systematic Review”. In: *Journal of Medical Internet Research* 23.3 (2021), e24387. DOI: 10.2196/24387. URL: https://doi.org/10.2196/24387.

[9] N. Bousbia et al. “Learners’ navigation behavior identification based on trace analysis”. In: *User Modeling and User-Adapted Interaction* 20.5 (2010), pp. 455–494. DOI: 10.1007/s11257-010-9081-5.

[10] A.G. Brady. “Uncovering Patterns in Process Data to Analyze Interactions and Learning Outcomes Within a Computer-Based Learning Environment”. In: *Research in Science Education* 54.1 (2024), pp. 83–100. DOI: 10.1007/s11165-023-10109-6.

[11] E. F. Buraimoh. “Predicting Student Success Using Student Engagement in the Online Component of a Blended-Learning Course”. Research-Report submitted in partial fulfilment for the degree of M.Sc (Coursework and Research Report) Computer Science. Master of Science Research-Report. Johannesburg, South Africa: The University of the Witwatersrand, Johannesburg, May 2021.

[12] H. Chai, T. Hu, and G. Niu. “How proactive personality promotes online learning performance? Mediating role of multidimensional learning engagement”. In: *Education and Information Technologies* 28.4 (2023), pp. 4795–4817. DOI: 10.1007/s10639-022-11319-7.

[13] Angela T. Chen et al. “A multi-faceted approach to characterizing user behavior and experience in a digital mental health intervention”. In: *Journal of Biomedical Informatics* 94 (2019), p. 103187. DOI: 10.1016/j.jbi.2019.103187. URL: https://doi.org/10.1016/j.jbi.2019.103187.

[14] E. S. Chen and J. J. Cimino. “Patterns of usage for a Web-based clinical information system”. In: *Stud Health Technol Inform* 107.Pt 1 (2004), pp. 18–22.

[15] M. Cocea and S. Weibelzahl. “Cross-System Validation of Engagement Prediction from Log Files”. In: *Creating New Learning Experiences on a Global Scale*. Ed. by E. Duval, R. Klamma, and M. Wolpers. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 14–25. ISBN: 978-3-540-75195-3.

[16] M. Cocea and S. Weibelzahl. “Disengagement detection in online learning: Validation studies and perspectives”. In: *IEEE Transactions on Learning Technologies* 4.2 (2011), pp. 114–124. DOI: 10.1109/TLT.2010.14.

[17] M. Cocea and S. Weibelzahl. “Log file analysis for disengagement detection in e-Learning environments”. In: *User Modeling and User-Adapted Interaction* 19.4 (2009), pp. 341–385. DOI: 10.1007/s11257-009-9065-5.

[18] O. Dermy, A. Roussanaly, and A. Boyer. “Using Behavioral Primitives to Model Students’ Digital Behavior”. In: *Procedia Computer Science*. Vol. 207. 2022, pp. 2444–2453. DOI: 10.1016/j.procs.2022.09.302.

[19] M. Ali Akber Dewan, Mahbub Murshed, and Fuhua Lin. “Engagement detection in online learning: a review”. In: *Smart Learning Environments* 6.1 (Dec. 2019). DOI: 10.1186/s40561-018-0080-z.

[20] A.A. Elrahman et al. “A Predictive Model for Student Performance in Classrooms using Student Interactions with an eTextbook”. In: *Information Sciences Letters* 12.1 (2023), pp. 9–12. DOI: 10.18576/isl/120102.

[21] S. Esnashari, L. Gardner, and M. Rehm. “Characterizing students’ behavior based on their participation in property course in New Zealand”. In: *Proceedings - 32nd IEEE International Conference on Advanced Information Networking and Applications Workshops, WAINA 2018*. Vol. 2018-Janua. 2018, pp. 6–11. ISBN: 9781538653944. DOI: 10.1109/WAINA.2018.00055.

[22] K. Etminani et al. “Web Usage Mining: Discovery of the users’ navigational patterns using SOM”. In: *2009 1st International Conference on Networked Digital Technologies, NDT 2009*. 2009, pp. 244–249. ISBN: 9781424446155. DOI: 10.1109/NDT.2009.5272158.

[23] J. Edmond Meku Fotso et al. "Algorithms for the Development of Deep Learning Models for Classification and Prediction of Behaviour in MOOCS." In: *Proceedings of 2020 IEEE Learning With MOOCS, LWMOOCS 2020*. 2020, pp. 180–184. ISBN: 9781728197289. DOI: 10.1109/LWMOOCS50143.2020.9234363.

[24] T. GopalaKrishnan and P. Sengottuvelan. "A hybrid PSO with Naïve Bayes classifier for disengagement detection in online learning". In: *Program* 50.2 (2016), pp. 215–224. DOI: 10.1108/PROG-07-2015-0047.

[25] S. Greiff et al. "Understanding students' performance in a computer-based assessment of complex problem solving: An analysis of behavioral data from computer-generated log files". In: *Computers in Human Behavior* 61 (2016), pp. 36–46. DOI: 10.1016/j.chb.2016.02.095.

[26] S. Hachmoud et al. "Analysis of students online learning behavior in a pedagogical model combining blended learning and competency based approach". In: *International Journal of Advanced Trends in Computer Science and Engineering* 8.6 (2019), pp. 3389–3395. DOI: 10.30534/ijatcse/2019/113862019.

[27] E. A. Hægermark, J. Nemeth, and A. Faxvaag. "How to Tell Whether Patients Engage and Use a Patient Portal - An Analysis of Five Functions". In: *Stud Health Technol Inform* 316 (Aug. 2024), pp. 1179–1183. DOI: 10.3233/SHTI240621.

[28] J.Y. Han et al. "Unraveling uses and effects of an interactive health communication system". In: *Journal of Broadcasting and Electronic Media* 53.1 (2009), pp. 112–133. DOI: 10.1080/08838150802643787.

[29] Q. He, F. Borgonovi, and M. Paccagnella. "Leveraging process data to assess adults' problem-solving skills: Using sequence mining to identify behavioral patterns across digital tasks". In: *Computers and Education* 166 (2021). DOI: 10.1016/j.compedu.2021.104170.

[30] Qingchun Hu, Yong Huang, and Liping Deng. "A framework to evaluate student engagement of programming course in blend learning environment". In: *ICCSE 2021 - IEEE 16th International Conference on Computer Science and Education*. Institute of Electrical and Electronics Engineers Inc., Aug. 2021, pp. 915–919. ISBN: 9781665414685. DOI: 10.1109/ICCSE51940.2021.9569553.

[31] K. Huber and M. Bannert. "Investigating learning processes through analysis of navigation behavior using log files". In: *Journal of Computing in Higher Education* 36.3 (2024), pp. 683–701. DOI: 10.1007/s12528-023-09372-3.

[32] Mark I. Hwang and Ron G. Thorn. "The Effect of User Engagement on System Success: A Meta-Analytical Integration of Research Findings". In: *Information & Management* 35.4 (1999), pp. 229–236. ISSN: 0378-7206. DOI: 10.1016/S0378-7206(98)00092-5. URL: https://www.sciencedirect.com/science/article/pii/S0378720698000925.

[33] Institute for Health Metrics and Evaluation. *Mental Health*. https://www.healthdata.org/research-analysis/health-risks-issues/mental-health. Accessed: 2024-10-08.

[34] E. Isohanni and M. Knobelsdorf. "Students' long-term engagement with the visualization tool VIP". In: *Proceedings of the 11th Koli Calling International Conference on Computing Education Research*. Koli Calling '11. Koli, Finland: Association for Computing Machinery, 2011, pp. 33–38. ISBN: 9781450310529. DOI: 10.1145/2094131.2094138. URL: https://doi.org/10.1145/2094131.2094138.

[35] Dr. Rozita Jamili oskouei, Mehrzad Askari, and P.R.P. Sajja. "Perceived Internet Usage Behaviours as Predictor to Outlier Detection in Students' Communities in Academic Environments". In: *Indian Journal of Science and Technology* 6 (July 2013), pp. 4923–4935. DOI: 10.17485/ijst/2013/v6i7.8.

[36] N. Kadoic and D. Oreski. "Analysis of student behavior and success based on logs in Moodle". In: *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2018 - Proceedings*. 2018, pp. 654–659. ISBN: 9789532330977. DOI: 10.23919/MIPRO.2018.8400123.

[37] L. Korzeniowski and K. Goczyla. "Landscape of Automated Log Analysis: A Systematic Literature Review and Mapping Study". In: *IEEE Access* 10 (2022), pp. 21892–21913. DOI: 10.1109/ACCESS.2022.3152549.

[38] Benson S. Ku et al. "Associations between Mental Health Shortage Areas and County-Level Suicide Rates among Adults Aged 25 and Older in the USA, 2010 to 2018". In: *General Hospital Psychiatry* 70 (2021), pp. 44–50. ISSN: 0163-8343. DOI: 10.1016/j.genhosppsych.2021.02.001. URL: https://www.sciencedirect.com/science/article/pii/S0163834321000219.

[39] Sanna Lehtimaki et al. "Evidence on Digital Mental Health Interventions for Adolescents and Young People: Systematic Overview". In: *JMIR Mental Health* 8.4 (2021), e25847. DOI: 10.2196/25847.

[40] K. Liang, D. Tu, and Y. Cai. "Using Process Data to Improve Classification Accuracy of Cognitive Diagnosis Model". In: *Multivariate Behavioral Research* 58.5 (2023), pp. 969–987. DOI: 10.1080/00273171.2022.2157788.

[41] Julia Löchner et al. "Digital Interventions in Mental Health: An Overview and Future Perspectives". In: *Internet Interventions* 40 (2025), p. 100824. DOI: 10.1016/j.invent.2025.100824.

[42] D. Logan. "A Comparative Measure of Student Engagement Using Log Files and a Student Self-report Survey". In: *Information Management*. Ed. by S. Li. Cham: Springer Nature Switzerland, 2024, pp. 321–331. ISBN: 978-3-031-64359-0. DOI: 10.1007/978-3-031-64359-0_26.

[43] A.V. Manjarres, L.G.M. Sandoval, and M.J.S. Suárez. "Data mining techniques applied in educational environments: Literature review". In: *Digital Education Review* 33 (2018), pp. 235–266.

[44] G. Mehta et al. "Enrichment of Student Performance model using Collaborative Methods". In: *2021 4th International Conference on Recent Developments in Control, Automation and Power Engineering, RDCAPE 2021*. 2021, pp. 195–198. ISBN: 9781665414296. DOI: 10.1109/RDCAPE52977.2021.9633499.

[45] Q. Nguyen, S. Thorne, and B. Rienties. "How do students engage with computer-based assessments: impact of study breaks on intertemporal engagement and pass rates". In: *Behaviormetrika* 45.2 (2018), pp. 597–614. DOI: 10.1007/s41237-018-0060-1.

[46] Claire de Oliveira et al. "The Role of Mental Health on Workplace Productivity: A Critical Review of the Literature". In: *Applied Health Economics and Health Policy* 21.2 (2023), pp. 167–193. DOI: 10.1007/s40258-022-00788-4.

[47] Y. Opanasenko, M. Pedaste, and L. A. Siiman. “Sequence Analysis as a Tool for Visualizing Students’ Patterns of Taking Tests in Interactive E-Book Environment”. In: *Innovative Technologies and Learning*. Ed. by Y. P. Cheng et al. Cham: Springer Nature Switzerland, 2024, pp. 120–128. ISBN: 978-3-031-65881-5.

[48] S. Y. Park, C. Nicksic Sigmon, and D. Boeldt. “A Framework for the Implementation of Digital Mental Health Interventions: The Importance of Feasibility and Acceptability Research”. In: *Cureus* 14 (Sept. 2022), e29329. DOI: 10.7759/cureus.29329.

[49] E. Pesare, T. Roselli, and V. Rossano. “Engagement in Social Learning: Detecting Engagement in Online Communities of Practice”. In: *Advances in Human Factors, Business Management, Training and Education*. Ed. by J. I. Kantola et al. Cham: Springer International Publishing, 2017, pp. 151–158. ISBN: 978-3-319-42070-7. DOI: 10.1007/978-3-319-42070-7_15.

[50] Q. Pham et al. “A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: Scoping review”. In: *JMIR mHealth and uHealth* 7.1 (2019). DOI: 10.2196/11941.

[51] L. Politi et al. “Use patterns of health information exchange through a multidimensional lens: Conceptual framework and empirical validation”. In: *Journal of Biomedical Informatics* 52 (2014), pp. 212–221. DOI: 10.1016/j.jbi.2014.07.003.

[52] S. Rathi, K. Kant Hiran, and S. Sakhare. “Affective state prediction of E-learner using SS-ROA based deep LSTM”. In: *Array* 19 (2023). DOI: 10.1016/j.array.2023.100315.

[53] RECONNECTED. *RECONNECTED Website*. https://reconnected-project.eu/. Accessed: 2026-07-22.

[54] F. Reinhold et al. “Process data from electronic textbooks indicate students’ classroom engagement”. In: *Learning and Individual Differences* 83-84 (2020). DOI: 10.1016/j.lindif.2020.101934.

[55] C. Romero et al. “Web usage mining for predicting final marks of students that use Moodle courses”. In: *Computer Applications in Engineering Education* 21.1 (2013), pp. 135–146. DOI: 10.1002/cae.20456.

[56] C.S. Sasikumar and A. Kumaravel. “Linear regression and support vector machine for classification of e-learning students’ engagement and performance”. In: *International Journal of Engineering and Advanced Technology* 8.6 (2019), pp. 2696–2700. DOI: 10.35940/ijeat.F8760.088619.

[57] C.S. Sasikumar and A. Kumaravel. “Supervised models for measuring performance at e-learning environment”. In: *International Journal of Innovative Technology and Exploring Engineering* 8.11 (2019), pp. 2792–2797. DOI: 10.35940/ijitee.K2279.0981119.

[58] A. Van Schalkwyk et al. “A Scoping Review of the Use of Log Data for Evaluating Mobile Apps: Exploring Implications for mHealth Apps”. In: *IEEE Access* 10 (2022), pp. 124805–124820. DOI: 10.1109/ACCESS.2022.3224038.

[59] A. Van Schalkwyk et al. “Contributing to monitoring and evaluation practices by conceptualizing a Valuable Insights from Log Analysis (VILA) Framework”. In: *2022 IEEE 28th International Conference on Engineering, Technology and Innovation, ICE/ITMC 2022 and 31st International Association for Management of Technology, IAMOT 2022 Joint Conference - Proceedings*. 2022. ISBN: 9781665488174. DOI: 10.1109/ICE/ITMC-IAMOT55089.2022.10033267.

[60] Priyanka Shirsat. "Developing Deep Neural Network for Learner Performance Prediction in EKhool Online Learning Platform". In: *Multimedia Research* 3.4 (Oct. 2020), pp. 24–31. DOI: 10.46253/j.mr.v3i4.a3.

[61] F. Sieverink et al. "Opening the black box of electronic health: Collecting, analyzing, and interpreting log data". In: *JMIR Research Protocols* 6.8 (2017). DOI: 10.2196/resprot.6452.

[62] P.V. Praveen Sundar and A.V. Kumar. "An Enhanced Disengagement Detection in Online Learning using Quasi Framework". In: *International Journal of Applied Engineering Research 0973-4562* 10 (June 2015), pp. 1298–1302.

[63] P.V. Praveen Sundar and A.V. Kumar. "Disengagement Detection in Online Learning Using Log File Analysis". In: *International Journal of Control Theory and Applications* 9 (Oct. 2016), pp. 295–301.

[64] P.V.P. Sundar and A.V.S. Kumar. "A novel disengagement detection strategy for online learning using quasi framework". In: *Souvenir of the 2015 IEEE International Advance Computing Conference, IACC 2015*. 2015, pp. 634–638. ISBN: 9781479980475. DOI: 10.1109/IADCC.2015.7154784.

[65] C.-Y. Ting and S. Phon-Amnuaisuk. "Log data approach to acquisition of optimal Bayesian learner model". In: *American Journal of Applied Sciences* 6.5 (2009), pp. 913–921. DOI: 10.3844/ajas.2009.913.921.

[66] G.D. Tosto et al. "Metrics for outpatient portal use based on log file analysis: Algorithm development". In: *Journal of Medical Internet Research* 22.6 (2020). DOI: 10.2196/16849.

[67] E. Ulitzsch, Q. He, and S. Pohl. "Innovations in Exploring Sequential Process Data". In: *Zeitschrift fur Psychologie / Journal of Psychology* 232.2 (2024), pp. 71–73. DOI: 10.1027/2151-2604/a000560.

[68] E. Ulitzsch, Q. He, and S. Pohl. "Using Sequence Mining Techniques for Understanding Incorrect Behavioral Patterns on Interactive Tasks". In: *Journal of Educational and Behavioral Statistics* 47.1 (2022), pp. 3–35. DOI: 10.3102/10769986211010467.

[69] A. Wakjira and S. Bhattacharya. "Predicting Student Engagement in the Online Learning Environment". In: *International Journal of Web-Based Learning and Teaching Technologies* 16.6 (2021). DOI: 10.4018/IJWLTT.287095.

[70] Q. Wang et al. "Examining adults' behavioral patterns in a sequence of problem solving tasks in technology-rich environments". In: *Computers in Human Behavior* 147 (2023). DOI: 10.1016/j.chb.2023.107852.

[71] Qin Wang and Amin Mousavi. "Which log variables significantly predict academic achievement? A systematic review and meta-analysis". In: *British Journal of Educational Technology* 54 (Oct. 2022). DOI: 10.1111/bjet.13282.

[72] World Health Organization. *Mental Health*. https://www.who.int/health-topics/mental-health. Accessed: 2026-07-22.

[73] H.Ç. Yavuz. "The effects of log data on students' performance | Bilgisayar ortamında kaydedilen madde yanıtlama verilerinin öğrenci performansına etkisi". In: *Journal of Measurement and Evaluation in Education and Psychology* 10.4 (2019), pp. 378–390. DOI: 10.21031/EPOD.564232.

[74] Jing Zhou et al. "Investigating response behavior through TF-IDF and Word2vec text analysis: A case study of PISA 2012 problem-solving process data". In: *Heliyon* 10.16 (Aug. 2024). ISSN: 24058440. DOI: 10.1016/j.heliyon.2024.e35945.

Appendix A

Engagement Metrics

A.1 Computation Procedures

Table A.1: Computation procedures for app engagement metrics.

Metric	**Computation**
Total time spent on the app	From `app.csv` sort by `app.username`, `app.timestamp`. A session *starts* when `app.event` $\in$ {"open", "resume"} and *ends* when `app.event = "pause"` for the same `app.username`. For each start–end pair, compute Δt = `app.timestamp`$_{\text{end}}$ − `app.timestamp`$_{\text{start}}$. For each `app.username`, sum all retained Δt values and report the total in minutes.
Average time spent in pages	Read `app_pages.csv` (fields `app_pages.username`, `app_pages.timestamp`, `app_pages.page_visited`) and label each row as a *page_visit*. From `app.csv`, keep only pause events (`app.event = "pause"`) with fields `app.username`, `app.timestamp`, and label as *pause*. Concatenate the two event streams into $username, timestamp, event\_type, page$ and sort by `username`, `timestamp`. For each *page_visit*, scan forward within the same `username` to the first subsequent event where $\Delta t = \text{next.timestamp} - \text{current.timestamp} \leq 30$ minutes and `event_type` $\in \{\text{page\_visit}, \text{pause}\}$; if such an event exists, record that Δt as the page duration, otherwise discard (idle). Aggregate by `username` and compute the mean of recorded durations; convert seconds to minutes and round.

Metric	Computation
Average time spent in active weeks	From `app_login.csv`, obtain each user's first login (`app_login.username`, `app_login.timestamp`; earliest by timestamp). From `app.csv`, keep only events with `app.event` $\in$ {"open", "resume", "pause"} and `app.timestamp` $\geq$ `app_login.timestamp` for the same `username`. Derive ISO fields `app.year`, `app.week` from `app.timestamp` and sort by `app.username`, `app.timestamp`. For each user and each $(\text{year}, \text{week})$ group, scan chronologically: whenever an *open/resume* is followed by the next *pause*, compute $\Delta t = \texttt{app.timestamp}_{\text{pause}} - \texttt{app.timestamp}_{\text{start}}$; if $0 < \Delta t \leq 30$ minutes, add to that week's total seconds. Weeks with `total_seconds` > 0 are *active*. The metric equals the average of weekly totals across active weeks only, converted to minutes and rounded.
Average duration of a session	Read `app.csv` and keep rows where `app.event` $\in$ {"open", "resume", "pause"}. Sort by `app.username`, `app.timestamp`. For each `app.username`, scan chronologically: whenever you encounter a start event (*open/resume*), find the next *pause*; compute $\Delta t = \texttt{app.timestamp}_{\text{pause}} - \texttt{app.timestamp}_{\text{start}}$. If $0 < \Delta t \leq 30$ minutes, record Δt (in seconds) as one session duration; otherwise discard. After processing all events for the user, compute the arithmetic mean of the recorded session durations and convert to minutes (rounded).
Average daily time	Read `app.csv` and keep rows where `app.event` $\in$ {"open", "resume", "pause"}. Sort by `app.username`, `app.timestamp`. For each `app.username`, scan chronologically: when a start event (*open/resume*) is followed by the next *pause*, compute $\Delta t = \texttt{app.timestamp}_{\text{pause}} - \texttt{app.timestamp}_{\text{start}}$. If $0 < \Delta t \leq 30$ minutes, add Δt (in seconds) to that day's data keyed by `date(app.timestamp_start).` After processing all events, compute `average_daily_seconds` and convert to minutes (rounded).

Metric	Computation
App opens / pauses / resumes	**App opens — Computation:** From app_count.csv, filter app_count.event = "open". **App pauses (closes) — Computation:** From app_count.csv, filter app_count.event = "pause". **App resumes — Computation:** From app_count.csv, filter app_count.event = "resume".
Daily average number of visits	Read app.csv; keep rows with app.event ∈ {"open","resume"}. Group by app.username, app.event_date and count rows per day. Then, per app.username, compute the mean of the per-day counts.
Total number of entries in the logs	For each file in {app_login.csv, app_notifications.csv, app_pages.csv, app_ratings.csv, app_records.csv, app_segments.csv, app.csv}, count rows per <file>.username (excluding nulls) and **sum** these counts across files for each user.
Number of ratings	Read app_ratings_count.csv and use the per-user count in app_ratings_count.count.
Number of records	Read app_records_count.csv (per-user, per-type counts in app_records_count.count keyed by app_records_count.recordtype).
Days of absence since first login	From app_login.csv, compute first_login_date = date(min(app_login.timestamp)) per app_login.username. From app.csv, take rows with app.event = "pause" and compute last_observed_date = date(max(app.timestamp)) per app.username. From app.csv, keep rows with app.event ∈ {"open","resume"}, derive app.date = date(app.timestamp), and collect the set active_dates per app.username. For each user with both dates present, build the inclusive date range [first_login_date, last_observed_date]; define absence_days = range \active_dates.

Metric	Computation
Temporal preference	From `app.csv`, derive `hour = hour(app.timestamp)` and map to `period` ∈ {`morning`, `afternoon`, `evening`, `night`} using the boundaries above. Count events by `(app.username, period)`, select the period with the maximum count per user, and map to `preference_id` via {`morning:1, afternoon:2, evening:3, night:4`}.
Average pages per session	From `app.csv`, keep `app.event` ∈ {open, resume, pause} and sort by `app.username`, `app.timestamp`. Reconstruct sessions as intervals [`session_start`, `session_end`] where `session_start = app.timestamp` at an open/resume and `session_end = timestamp of next pause`. From `app_pages.csv`, count rows where `app_pages.username = session.username` and `app_pages.timestamp` ∈ [`session_start`, `session_end`]; this yields `pages_in_session`. Average `pages_in_session` across a user's sessions.
Average unique pages per session	Build sessions from `app.csv`. For each session, filter `app_pages.csv` to the same `username` and timestamps within [`session_start`, `session_end`], then compute `nunique(app_pages.page_visited)` to obtain `unique_pages_in_session`. Average `unique_pages_in_session` across the user's sessions.
% ratings without notification	Read `app_ratings.csv`. For each `app_ratings.username`, compute `total_ratings = count(*)` and `ratings_without_notif = count(source = "")`. Report $$\%\text{ without notif} = 100 \times \frac{\texttt{ratings\_without\_notif}}{\texttt{total\_ratings}}$$ as `app_ratings_without_notif_percent` (rounded).

Metric	Computation
% ratings via OS notification	Read `app_ratings.csv`. For each `app_ratings.username`, compute `total_ratings = count(*)` and `ratings_device = count(source = "device")`. Report $$\text{\% via OS notif} = 100 \times \frac{\texttt{ratings\_device}}{\texttt{total\_ratings}}$$ as `app_ratings_system_notif_percent` (rounded).
Average length of record answers	Read `app_records.csv`; keep rows with `app_records.recordtype` $\in$ {"food","thought"}. Parse `app_records.content` (JSON) and extract the text at key `recordtype`; call it `response`. Exclude missing/empty responses, compute `response_length = len(response)`, then group by `app_records.username`, `app_records.recordtype` and take the mean.
Longest consecutive stretch of active weekly usage	From `app_login.csv`, compute `first_login_date = date(min(app_login.timestamp))` per `app_login.username`. From `app.csv`, compute `last_observed_date = date(max(app.timestamp))` per `app.username` using `app.event = "pause"`. From `app.csv`, keep `app.event` $\in$ {"open","resume"} and derive `app.date = date(app.timestamp)`. For each user, restrict to `first_login_date` $\leq$ `app.date` $\leq$ `last_observed_date` and map each `app.date` to (ISO year, ISO week) $\rightarrow$ set `active_weeks`. Build the ordered inclusive ISO-week range from `first_login_date` to `last_observed_date`. Scan the range; increment streak when week $\in$ `active_weeks`, else reset; keep the maximum.

Metric	Computation
% of weeks active after first login	From `app_login.csv`, `first_login_date = date(min(app_login.timestamp))` per user. From `app.csv`, `last_observed_date = date(max(app.timestamp))` where `app.event = "pause"` per user. From `app.csv`, keep `app.event` $\in$ {"open","resume"}, derive `app.date = date(app.timestamp)`, restrict to the window, and map to `active_weeks = set of (ISO year, ISO week)`. Construct `all_weeks` as the set of ISO weeks in the inclusive date range (iterate week-by-week). Let `total_weeks = |all_weeks|`, `active_weeks_count = |active_weeks|`, and compute $$\texttt{app\_act\_weeks\_percent} = 100 \times \frac{\texttt{active\_weeks\_count}}{\texttt{total\_weeks}}.$$

Table A.2: Computation procedures for portal engagement metrics.

Metric	Computation
Total time spent on the portal	Read `patient_logs.csv`; sort by `patient_logs.username`, `patient_logs.serverTimeStamp`. Track `current_login_time[user]` and `last_event_time[user]`. On `loginW`: if the user is already logged in, close the prior session as Δt = `last_event_time` − `current_login_time` (if $\Delta t > 0$), then set `current_login_time = serverTimeStamp`. On `logoutW`: if logged in, close the session as Δt = `serverTimeStamp` − `current_login_time` (if $\Delta t > 0$), then clear `current_login_time`. After processing all rows: if a user remains logged in, close the session at `last_event_time`. Sum all retained durations per `patient_logs.username` and convert to minutes.

Metric	Computation
Total time spent on modules	Read `patient_page_timings.csv`; group by `patient_page_timings.username`, sum `elapsed_time_milliseconds` and convert to minutes.
Average time spent in pages	Read `patient_page_timings.csv`. Group by `patient_page_timings.username` and take the **mean** of `elapsed_time_milliseconds` and convert to minutes.
Average time spent in active weeks	Read `patient_logs.csv`; sort by `patient_logs.username`, `patient_logs.serverTimeStamp`. Derive `patient_logs.week` as `to_period("W").start_time`. For each `patient_logs.username`, iterate chronologically: on `loginW`, hold `login_time` and `week`; find session end; accumulate (`session_end` − `login_time`) (in seconds) into `weekly_time[week]`. After scanning, compute the average of the `weekly totals` across keys in `weekly_time` and convert to minutes.
Average duration of a session	Read `patient_logs.csv`; sort by `patient_logs.username`, `patient_logs.serverTimeStamp`. For each user, scan chronologically to build session intervals [start = "loginW", end]; compute `duration_sec = end - start` and collect all session durations. Compute the arithmetic mean per user and convert to minutes.
Average daily time	Read `patient_logs.csv`; sort by `patient_logs.username`, `patient_logs.serverTimeStamp`. Reconstruct sessions. For each session, assign its duration to `date(start)` where `start` is the `"loginW"` timestamp; aggregate per-day totals per user. For each user, compute `avg_daily_seconds = (sum of day totals) / (number of active days)` and convert to minutes.

Metric	Computation
Average time to complete a module	Read `patient_modules.csv`. Group by `patient_modules.username` and take the **mean** of `hoursInModule`.
Number of logins	Read `patient_login_count.csv`; use `patient_login_count.logins` as the per-user total.
Daily average number of visits	Read `patient_login_logout.csv`; keep rows where `patient_login_logout.operationType = "loginW"`. Derive `date = date(patient_login_logout.serverTimeStamp)`. Count logins per `(username, date)`, then per `username` compute the **mean** of those daily counts.
Total number of entries in the logs	Read `patient_logs.csv`; group by `patient_logs.username` and compute **size()** (row count).
Average module pages per session	From `patient_logs.csv`, sort by `patient_logs.username`, `patient_logs.serverTimeStamp`, and reconstruct sessions $[\text{start}, \text{end}]$. From `patient_page_timings.csv`, parse `enter_time` and `exit_time`. For each session, count rows where `username` matches and `enter_time` ∈ `[start, end]` → `pages_completed`. Per `username`, average `pages_completed` over all sessions.
Average unique module pages per session	Reconstruct sessions from `patient_logs.csv`. For each session, filter `patient_page_timings` to the same `username` with `enter_time` ∈ `[start, end]`; drop duplicates on `[module, page]` and count → `unique_pages_completed`. Per `username`, average `unique_pages_completed` across sessions.
Number of modules completed	Read `patient_modules.csv`; filter where `patient_modules.statusId = "completed"`. Group by `patient_modules.username` and count rows.

Metric	Computation
% of modules completed	From `patient_modules.csv`, compute `total_modules = count(*)` per `patient_modules.username`. Filter where `statusId = "completed"`, compute `completed_modules` per user. Join and compute $$\text{completion\_percentage} = 100 \times \frac{\text{completed\_modules}}{\text{total\_modules}}.$$
Days of absence since first login	Read `patient_login_logout.csv`. Derive `date = date(serverTimeStamp)`. Filter to `operationType = "loginW"`. For each `patient_login_logout.username`, take the unique login dates; let `first = min(date)` and `last = max(date)`. Compute `total_days = (last - first) + 1`, `login_days = |unique login dates|`, and `non_login_days = total_days - login_days`.
Longest consecutive stretch of active weekly usage	Read `patient_login_logout.csv`; keep `operationType = "loginW"`; derive `week = to_period("W").start_time` from `serverTimeStamp`. For each `username`, take the sorted unique week values and scan them in order: if the difference from the previous week is exactly 7 days, extend the current streak; otherwise, reset. Track the maximum.
% of weeks active after first login	Read `patient_login_logout.csv`; keep `operationType = "loginW"`; derive `week = to_period("W").start_time`. For each `username`, get sorted unique week values; let `first_week = min(week)` and `last_week = max(week)`. Compute `total_possible_weeks = ((last_week - first_week).days // 7) + 1` and `active_weeks = |unique week values|`. Report $$\text{portal\_act\_weeks\_percent} = 100 \times \frac{\text{active\_weeks}}{\text{total\_possible\_weeks}}.$$

Metric	**Computation**
Temporal preference	Read `patient_login_logout.csv`; derive `hour = hour(serverTimeStamp)` and map to `interval_id` $\in$ `{1,2,3,4}` using the boundaries above. Restrict to `operationType = "loginW"` and count logins by `(username, interval_id)`. For each `username`, select the `interval_id` with the highest `login_count`.

Appendix B

Interaction Data Analysis

B.1 Portal Modules Coverage

This section presents a visual representation of the percentage of pages visited and exercises completed for each module in the web portal.

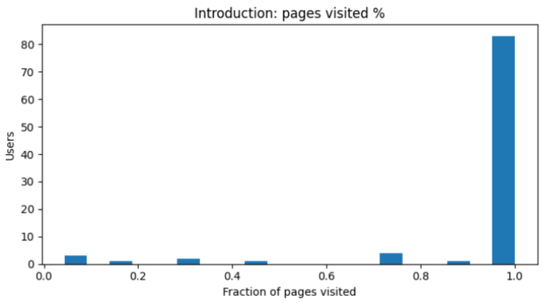


Figure B.1: Distribution of the fraction of pages visited per user in the "Introduction" module.

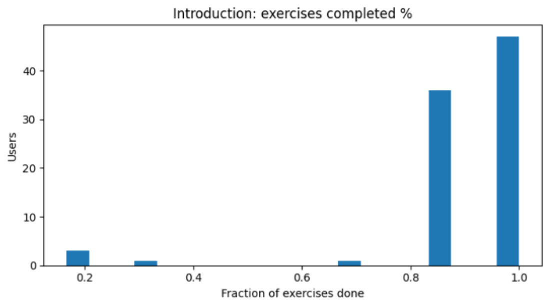


Figure B.2: Distribution of the fraction of exercises completed per user in the "Introduction" module.

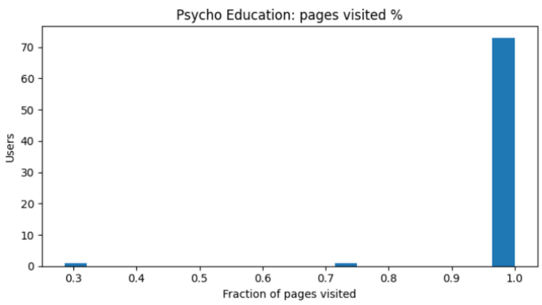


Figure B.3: Distribution of the fraction of pages visited per user in the "Psycho Education" module.

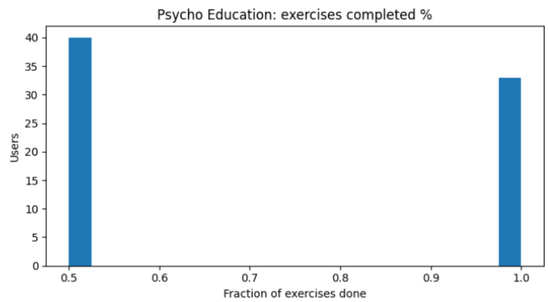


Figure B.4: Distribution of the fraction of exercises completed per user in the "Psycho Education" module.

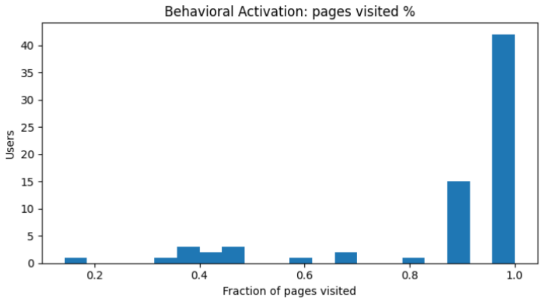


Figure B.5: Distribution of the fraction of pages visited per user in the "Behavioral Activation" module.

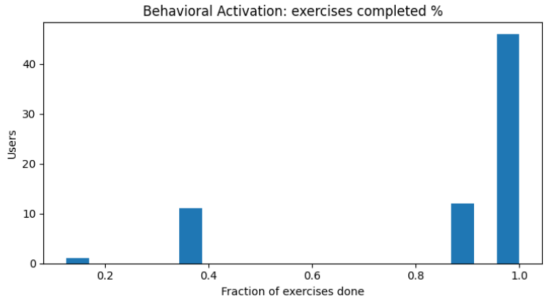


Figure B.6: Distribution of the fraction of exercises completed per user in the "Behavioral Activation" module.