U.PORTO

DOUTORAMENTO

CIÊNCIAS BIOMÉDICAS

Training in Laboratory Animal Science: new perspectives on practical competence achievement with rodents

Andreia Filipa Fontes da Costa

D

2022

U.PORTO

Andreia Filipa Fontes da Costa

Training in Laboratory Animal Science: new perspectives on practical competence achievement with rodents

D.ICBAS **2022**

Training in Laboratory Animal Science: new perspectives on practical competence achievement with rodents

Andreia Filipa Fontes da Costa

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR

Andreia Filipa Fontes Da Costa

Training in Laboratory Animal Science: New perspectives on practical competence achievement with rodents

Tese de Candidatura ao grau de Doutor em Ciências Biomédicas
do Programa Doutoral da Universidade do Porto, Instituto de Ciências Biomédicas de Abel Salazar

Orientador – Doutora Ingrid Anna Sofia Olsson
Categoria – Investigador Principal
Afiliação – Categoria – Investigador Principal
Afiliação – Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto / Instituto de Biologia Molecular e Celular

Co-orientador – Doutor Manuel João Tavares Mendes Costa
Categoria – Professor Associado
Afiliação – Escola de Medicina, Universidade do Minho

Co-orientadora – Doutora Maria Salomé Custódio Gomes
Categoria – Professora Associada
Afiliação – Instituto de Ciências Biomédicas de Abel Salazar / Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto / Instituto de Biologia Molecular e Celular

This thesis includes the following papers:

Published

Costa A, Costa A, Olsson IAS. Students' acceptance of e-learning approaches in laboratory animal science training. Laboratory Animals. 2019;0:1–11.

Costa A, Lamas S, Costa MJ, et al. The assessment of researchers' competence in experimental procedures with laboratory animals: A three-step methodology to develop a global rating scale. Laboratory Animals. 2021;55(5):463-71.

Submitted

Costa A, Lamas S, Correia MR, et al. An Objective Structured Laboratory Animal Science Examination (OSLASE) to ensure researcher's professional competence in LAS
Submitted to Special edition of Laboratory Animals Journal on Education.

First author contribution to the listed papers

The first author contributed to the study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

The papers listed above, integrating this thesis were or will not be included in other thesis or similar work.

AGRADECIMENTOS

AGRADECIMENTOS

Esta tese é o resultado de vários anos de trabalho que só foi possível graças às pessoas que cruzaram o meu caminho e me ajudaram das mais diversas formas. A algumas dessas pessoas, gostaria de expressar a minha gratidão.

À Doutora Anna Olsson, minha orientadora, agradeço a orientação e tudo o que me ensinou desde que cheguei ao laboratório. Pela liberdade que sempre me deu para expor as minhas ideias, para experimentar e corrigir, se assim fosse preciso. Pelas reuniões e conversas que tivemos. Pela paciência, apoio e pela confiança que depositou em mim.

Ao Professor Doutor Manuel João Costa, meu co-orientador, um mentor e elemento decisivo em vários momentos deste percurso. Obrigada pelas discussões científicas, pela critica construtiva, por me ajudar a abrir horizontes, pelos conselhos e pela palavra de ânimo que sempre teve.

A Professora Doutora Salomé Gomes, minha co-orientadora, por aceitar abraçar este projeto, pelo apoio constante, pela disponibilidade e amabilidade que sempre demostrou e pela contribuição para que este trabalho chegasse a bom porto.

À Sofia Lamas, que esteve presente em grande parte do trabalho prático apresentado nesta tese. Agradeço a amizade que criamos desde o início, o companheirismo e a “equipa” que as duas conseguimos criar.

Ao laboratório 211, a todos aqueles cuja partilha e reflexão de ideias contribuíram para este trabalho, em especial à Ana Magalhães, à Teresa Summavielle, à Joana Bravo, à Marta Silva pelo incentivo, sobretudo, nesta reta final. Às minha colegas de gabinete que me apoiaram nos momentos mais críticos de trabalho para que eu conseguisse avançar sem desmotivar. Aos demais colegas do instituto que direta ou indiretamente ajudaram na realização deste projeto. Ao grupo de trabalho da ETPLAS, pelo espírito de companheirismo, pelo muito que partilhamos e apendemos em conjunto.

Ao meu Pai, que percebi que esteve sempre comigo na maneira como vivi os desafios, os contratempos e as conquistas. Por tudo que me ensinou e pelo exemplo que me continua a inspirar. À minha mãe, pelo seu amor, dedicação e força, a minha maior apoiante. À minha irmã, a minha pessoa, terapeuta, amiga, conselheira. Partilhamos muitas aventuras, e esta minha, não foi diferente. À minha avó de coração, pela força que me deu sem o saber, pelo mimo em forma de maças assadas e leite creme. À Nonozeca, que chegou quase no fim deste percurso, mas a tempo de me motivar com melhores abraços e “águs” do mundo. Ao Pedro, pela amizade, pelo apoio, pela boa disposição e por estar sempre pronto ajudar. À minha tia Celta por me conhecer e compreender e me dizer aquilo que sempre preciso ouvir. E a toda a minha restante família, que sempre me apoiou.

À Ana, minha “gémea”, pelo apoio incondicional e por viver comigo as minhas vitórias e derrotas, às vezes até mais que eu, e por nunca me deixar baixar os braços.
À Catarina pelos anos de amizade e por saber que sempre estará ao meu lado no caminho.
A todos meus amigos.

'DOCENDO DISCIMUS'

Latin proverb
| We learn by teaching |

ABSTRACT

ABSTRACT

Researchers and all staff using animals to perform experiments must undergo training in order to develop the appropriate skills and competence. This appears not only as a legal European and national requirement, but also as a way to guarantee the 3Rs application in practice. Directive 2010/63/EU establishes in article 23 that “the staff shall be adequately educated and trained” and “supervised in the performance of their tasks until they have demonstrated the requisite competence”. Despite this need, there is a lack of research dedicated to teaching and assessment in Laboratory Animal Science (LAS).

In this context, the present thesis aims to study two of the most relevant topics that laboratory animal science education faces presently: on the one hand, the use of e-learning and, on the other hand, assessment and competence evaluation.

The use of e-learning has grown and it now plays an unquestionable role as a teaching tool, particularly in higher education. By providing flexible learning, it is seen as an adequate solution to help its users in terms of time, place and pace of learning. This advantage can be particularly useful in LAS training with intensive and demanding courses usually with a short duration (3-10 days). Also, e-learning allows that biomedical institutions with a regular need of new researchers’ training, implement strategies with a constant offer of standardized quality training contents. Nevertheless, the acceptance of e-learning by trainees and how they perceive its introduction in teaching is still unexplored. Also, no specific and valid instruments adjusted to the LAS context exists to measure the users’ acceptance. For this purpose, the Questionnaire of E-learning Acceptance (QELA), a 32-item accordance Likert-type scale comprising five subscales was developed to explore the following dimensions: participants’ perception of e-learning, satisfaction with organization and contents, perception of e-learning relevance for the time management, and its influence for practical training. The questionnaires were administrated to participants of Introductory courses (former FELASA category B) with all the theoretical contents online, and to the Advanced courses (former FELASA category C) with a combined and complementary theoretical approach of e-learning and classroom lectures. 229 respondents of 15 courses participated in this study: 90 participants of 9 introductory courses and 139 participants of 6 advanced courses.

Assessment is part of all the educational process, including in LAS. Here, the lack of researchers’ practical competence may result an unacceptably low standard of professional performance when using animals. We addressed the topic of competence assessment in two sequential studies: first, we proposed a methodology to develop a Global Rating Scale (GRS) to assess practical competence; second, we implemented an Objective Structured Laboratory Animal Science Examination (OSLASE). To analyse the implementation of OSLASE as a practical assessment strategy, the interrater reliability (IRR) was investigated and focus group interviews were also conducted to access trainees' acceptability.

The findings of these studies evidenced that e-learning is well accepted by researchers when integrated in LAS courses to deliver theoretical training. High levels of acceptance were shown, for both the introductory and the advanced training. Participants considered e-learning enjoyable, useful, and effective in this specific context, and sociodemographic, academic, and professional factors did not affect trainees' e-learning acceptance.

Regarding LAS assessment we have set out a 3 steps method, that can be used to create the instruments to assess proficiency in laboratory animal procedures. We propose the following sequential phases: (A) Item Development; (B) Scale Development; and (C) Piloting of the Scale. The process was coordinated by a core team with expertise in teaching LAS and in biomedical teaching methods, which are crucial topics to consider when developing an assessment instrument. Thus, this team input and dynamic is essential for the success of the methodology proposed.

Interrater agreement analysis of the OSLASE implementation revealed moderate to good values of weighted kappa and acceptable percentages of agreement. Trainees recognized that OSLASE promoted and guided their learning, increased their engagement and motivation during practical sessions, as well as served as a confirmation that relevant skills had been learned.

The studies presented in this thesis aimed to clarify current practical issues related with the implementation of strategies to deliver training and assess competence, thus, their findings are directly applicable to the LAS educational context. These studies bring insights with special focus on the delivery of knowledge and on assessment methods to support training, tutors and institutions, as well as in promoting training and ensuring competence in the work with laboratory animals.

KEYWORDS

Education and Training in Laboratory Animal Science, E-learning, Performance Assessment, Competence and Proficiency, Objective Structured Laboratory Animal Science Examination (OSLASE)

RESUMO

Treino em Ciências de Animais de Laboratório:

Novas perspetivas no desenvolvimento de competências práticas com roedores

RESUMO

Investigadores e todo o *staff* que usa animais para realizar experimentação, deve receber formação para desenvolver as capacidades adequadas e competência. Esta necessidade surge não só como um requisito legal europeu e nacional, mas também como forma de garantir a aplicação dos 3Rs na prática. A Diretiva 2010/63/UE estabelece no artigo 23.º que "o *staff* deve ser adequadamente formado e treinado" e "supervisionado no desempenho das suas funções até que demonstre a competência exigida". Apesar deste requisito, os estudos dedicados ao ensino e avaliação em Ciências de Animais de Laboratório (CAL) são escassos. Neste contexto, a presente tese visa estudar dois dos mais relevantes tópicos que o ensino de Ciências de Animais de Laboratório enfrenta atualmente: por um lado, o uso de e-learning e, por outro, a avaliação e verificação de competências.

A utilização do e-learning tem vindo a crescer, e hoje desempenha um papel inquestionável como ferramenta de ensino, nomeadamente, no ensino superior. Ao permitir uma aprendizagem flexível, é visto como uma solução adequada para ajudar os seus utilizadores no que respeita à gestão do tempo, local e ritmo da aprendizagem. Esta vantagem pode ser particularmente útil no treino em CAL, caracterizado por cursos intensivos e exigentes, geralmente de curta duração (3-10 dias). Além disso, o e-learning permite que instituições biomédicas com necessidade regular de formação de novos investigadores, implementem estratégias com uma oferta constante de conteúdos formativos de qualidade uniformizados. No entanto, a aceitação do e-learning pelos formandos e a forma como estes percecionam a sua utilização no ensino não foi ainda explorada. Além disso, não existem instrumentos específicos e válidos ajustados ao contexto de CAL para medir a aceitação dos utilizadores. Para este fim, foi desenvolvido o Questionário de Aceitação do E-learning (QAEL), de 32 itens, com uma escala de resposta de tipo Likert de concordância, composta por cinco subescalas com o objetivo de explorar as seguintes dimensões: perceção dos participantes sobre o e-learning, satisfação com organização e conteúdos, perceção da relevância do e-learning para a gestão do tempo, e a influência do e-learning para a formação prática. Os questionários foram aplicados aos participantes dos cursos Introdutórios (antigo FELASA categoria B), com todos os conteúdos teóricos online, e aos cursos Avançados (antigo FELASA categoria C) com uma abordagem teórica combinada e complementar de e-learning e aulas presenciais. Neste

estudo participaram 229 formandos de 15 cursos: 90 participantes de 9 cursos introdutórios e 139 participantes de 6 cursos avançados.

A avaliação é parte relevante em todos os processos educativos, incluindo em CAL. Em particular, a falta de competência prática dos investigadores pode resultar num padrão de desempenho profissional inaceitavelmente baixo no uso de animais. Segundo esta linha, abordamos o tema da avaliação de competência em dois estudos sequenciais: primeiro, propusemos uma metodologia para desenvolver uma Escala Avaliação Global (EAG) para avaliar a competência prática; em segundo lugar, implementamos um Exame em Ciência de Animais de Laboratório Objetivo Estruturado (ECALOE). Para analisar a implementação do ECALOE como uma estratégia de avaliação prática, a confiabilidade interobservador (CIO) foi investigada e entrevistas com grupos focais foram realizadas para avaliar a aceitação dos formandos em relação da ECALOE.

Os resultados destes estudos evidenciaram que o e-learning é bem aceite pelos investigadores quando integrado nos cursos CAL para administrar formação teórica. Verificaram-se níveis elevados de aceitação, tanto para formações introdutórias como para as avançadas. Os participantes consideraram o e-learning agradável, útil e eficaz neste contexto específico, sendo que fatores sociodemográficos, académicos e profissionais não afetaram a sua aceitação do e-learning. Em relação à avaliação em CAL, estabelecemos um método com 3 etapas, que pode ser usado para criar instrumentos para avaliar a proficiência em procedimentos em animais de laboratório. Assim, propomos as seguintes fases sequenciais: (A) Desenvolvimento de Itens; (B) Desenvolvimento de Escala; e (C) Estudo Piloto da Escala. O processo foi coordenado por uma equipa com experiência no ensino de CAL e em métodos de ensino biomédico, que são tópicos cruciais a ser considerados no desenvolvimento de um instrumento de avaliação. Assim, a dinâmica e participação desta equipa é essencial para o sucesso da metodologia proposta.

A análise de concordância entre avaliadores da implementação do ECALOE revelou valores de kappa ponderado de moderados a bons e percentagens de concordância aceitáveis. Os formandos reconheceram que o ECALOE promoveu e orientou a sua aprendizagem, aumentou o seu envolvimento e motivação durante as sessões práticas, bem como, permitiu a confirmação de que as competências relevantes foram aprendidas.

Os estudos apresentados nesta tese visaram esclarecer questões práticas atuais, relacionadas com a implementação de estratégias para administrar formação e avaliar competência, sendo os seus resultados diretamente aplicáveis ao contexto de ensino em CAL. Estes estudos trazem novas perspetivas com especial foco na transmissão de conhecimento teórico e nos métodos de avaliação, e os seus resultados visam apoiar a formação, tutores e instituições, bem como, promover o treino e assegurar competência no trabalho com animais de laboratório.

PALAVRAS CHAVE
Ensino e Treino em Ciências de Animais de Laboratório, E-learning, Avaliação de Desempenho, Competência e Proficiência, Exame em Ciência de Animais de Laboratório Objetivo Estruturado (ECALOE).

TABLE OF CONTENTS

TABLE OF CONTENTS

ABBREVIATIONS

ABBREVIATIONS

A

AAALAC - American Association for Accreditation of Laboratory Animal Care

ABIM - American Board of Internal Medicine

ASPA - Animals in Scientific Procedures Act

AWA - Animal Welfare Act

C

CBD - Case-based Discussion

CBEL - Case-based E-learning

CBL - Case-based Learning

CEC - Clinical Encounters Cards

CEX - Clinical Evaluation Exercise

CLs - Checklists

CMS - Course Management System

CPD - Continuing Professional Development

CWS - Clinical Work Samples

D

DOPS - Direct Observation of Procedural Skills / Directly Observed Practical Skills

DMV - Doctor of Veterinary Medicine

E

EEC - European Economic Community (abbreviated in the context of Directive 86/609/EEC)

EQs - Essay Questions

ERE - Emergency Remote Education

ETPLAS - Education and Training Platform for Laboratory Animal Science

ETS No. 123 - European Convention for the Protection of Vertebrate Animals Used for Experimental and Other Scientific Purposes

EWG - Expert Working Group

F

FELASA - Federation of European Laboratory Animal Science Associations

G

GRSs – Global Rating Scales

I

ISLAA - The Improved Standards for Laboratory Animals Act

IRR - Interrater Reliability

K

Kw - Weighted Kappa

L

LAS – Laboratory Animal Science

LMS - Learning Management System

M

MCQs - Multiple Choice Questions

MEQs - Multiple Essay Questions

miniCEX - Mini Clinical Evaluation Exercise

MSF – Multi-Source Feedback

N

NACWO - Named Animal Care and Welfare Officer

NVS - Named Veterinary Surgeon

O

OSATS – Objective Structured Assessment of Technical Skills

OSCE - Objective Structured Clinical Examination

OSLASE - Objective Structured Laboratory Animal Science Examination

OSLER - Objective Structured Long case

OSPRE/OSPE - Objective Structured Practical Exam

OSPVE - Objective Structured Practical Veterinary Exam

P

PACES - Practical Assessment of Clinical Examination Skills

PIL - Personal Licence

S

SPs - Standardized Patients' Simulations

V

VLE - Virtual Learning Environment

THESIS LAYOUT

THESIS LAYOUT

Chapter 1 presents the rationale and main aims of this thesis.

Chapter 2 presents a general introduction to the thesis work, in three sub-chapters. The first sub-chapter describes the history and background associated with education and training in Laboratory Animal Sciences (LAS), the second sub-chapter reports the LAS training courses and educational approaches used in the field and in similar educational areas, namely veterinary medicine, and the third one contextualizes the topic of competence assessment in health sciences education.

Chapter 3 presents the materials and methods used to develop the studies presented in this thesis.

Chapter 4 displays the research work on LAS students' acceptance of introducing e-learning approaches in their training. This chapter is presented as a final original paper published in *Laboratory Animals* Vol. 54(5) in 2019 (Costa et al. 2019).

Chapter 5 displays the research work on the methodology to develop a global rating scale for the assessment of researchers' competence in experimental procedures with laboratory animals. This chapter is presented as a final original paper published in *Laboratory Animals* Vol. 55(5) in 2021 (Costa et al. 2021).

Chapter 6 presents the research work exploring the implementation of an Objective Structured Laboratory Animal Science Exam (OSLASE) to assess researchers' competence. This chapter is presented in the form of the submitted paper to a special issue of *Laboratory Animals* on the topic of Education in Laboratory Animal Science and the 3Rs.

Chapter 7 encompasses a general discussion, in which the major findings of the research work are debated in light of the current literature and state of the art in LAS education.

Chapter 8 presents the thesis final remarks and future directions.

Appendices

I. Questionnaires administered to conduct the study reported in chapter 4, Students' acceptance of e-learning approaches in Laboratory Animal Science Training.

II. Global rating scales developed in the context of the work reported in chapter 5, to assess the competence of researchers in handling and performing substance administrations in laboratory rodents (Costa et al.). Published at the Open Repository of the University of Porto.

III. Published abstract: Poster communication “The development of a high stakes OSCE to enhance the assessment of practical competence in Laboratory Animal Science Courses” at Annual European Conference on Assessment in Medical Education ‘Technology Enabled Assessment’ (EACCME) 2018 (Poster12, Costa et al. Braga, Portugal 2018).

IV. Published abstract: Oral communication “Laboratory Animal Science online training through the eyes of the learner: perceptions of gains” at 14th FELASA Congress 2019 (OA4S2, Costa et al. Lab Anim 2019, 53(1S) pp.31).

V. Published abstract: Workshop organization “Are you skilled enough to work with animals? Measuring competencies using OSCE stations” at 14th FELASA Congress 2019 (OA12W1, Costa et al. Lab Anim 2019, 53 (1S) pp. 41).

VI. Published manuscript: “An Objective Structured Laboratory Animal Science Examination (OSLASE) to ensure researchers’ professional competence in laboratory animal science“, Costa et al. Lab Anim 2022, 0(0) 1- 11.

1. BACKGROUND AND AIMS

BACKGROUND AND AIMS

Adequate education in Laboratory Animal Science (LAS) and species-specific training are compulsory for those who conduct research and perform experimental procedures with laboratory animals, being a prerequisite for good animal welfare.

New and challenging requirements for education and training in the field of LAS were introduced in the recent decades, particularly promoted by the revised European legislation, together with the growing LAS research and consequent development of the field.

Despite these changes, our understanding on how education and training in LAS should be promoted, delivered and verified is still limited. Moreover, the interest in the design and implementation of innovative assessment methods is increasing noticeably. Besides that, the evaluation of competencies is legally required however
the lack of research in competence assessment topic in Laboratory Animals is recognized by the scientific community.

The overall objective of this thesis is to contribute: to improve delivery by exploring the impact of e-learning in the courses and develop assessment strategies in LAS education.

Two major aims were established for the work presented in this thesis:

1. To understand and clarify the trainees' perception and acceptance of using e-learning to deliver theoretical knowledge in LAS courses (Chapter 4).

2. To develop assessment instruments that enable measuring researchers' competence (Chapter 5).

3. To analyze the implementation of OSLASE – Objective Structured Laboratory Animal Science Exam in a training context (Chapter 6).

2. INTRODUCTION

INTRODUCTION

2.1 | EDUCATION AND TRAINING IN LABORATORY ANIMAL SCIENCE: BACKGROUND AND HISTORY

2.1.1 | OVERVIEW: EDUCATION AND TRAINING & LEGAL AND REGULATORY FRAMEWORK

The use of animals for teaching and research purposes began many hundred years ago. Scientific experiments involving animals are reported since Aristotle, Galen, and Vesalius (1, 2). The assumption that conducting experiments with living animals requires training to guarantee the welfare of the animals involved, emerged with the first efforts to regulate the animal experimentation.

The United Kingdom pioneered the issuing of legislation protecting animals used for scientific purposes. The Cruelty to Animals Act was enacted in 1876, following debate between researchers using animals and animal' protectionists. This legislation led to the creation of a central governing body dedicated to the review and approval of experiments using animals (3). United Kingdom example was followed decades later by several European countries. Since the Cruelty to Animals Act publication, movements emerged worldwide lead by antivivisection animal societies, and governmental and non-governmental animal welfare organizations, which influenced the actual societal vision on this matter. These movements also influenced the development of laws and regulations, as well as the establishment of standards. This fact shaped the development of the regulatory panorama and society posture on how animals in biomedical research can be used to pursue knowledge, and therefore our demands in terms of the education of personnel working with animals in science (2).

The directors of the Laboratory Animal Bureau formed in the UK (1947), R. E. Gloverand and W. Lane-Petter, reported concerns with the quality of animal care and with the scientific outputs using animals, together with the lack of a standardized education system in the field. Consequently, the laboratory promoted in April of 1948, the first of several conferences for animal care personnel held at the Royal Veterinary College in London. The first organizational meeting for the new Animal Technician Association (ATA), renamed several years later as Institute of Animal Technology [IAT], ran in the UK in August 1949 (2).

In the USA, the first events related with laboratory animals occurred in 1965, when researchers and veterinarians organized the American Association for Accreditation of Laboratory Animal Care (AAALAC), as a private and nonprofit organization. AAALAC started accrediting organizations across the country and issuing the majority of the standards to use animals in research, including guidelines for teaching (2, 4). Twenty years later, in 1985, the main federal law in USA, regarding the laboratory animal protection – the Animal Welfare Act (AWA), was considerably revised, after the report of numerous documented violations to the legislation. The Improved Standards for Laboratory Animals Act (ISLAA) was amended to the AWA, intended to minimize the pain and distress of animals in laboratory (5). The amendments provided major changes in the regulation regarding key aspects in the field, namely the training of personnel working with animals and the responsibility of the committees to guarantee that institutions were properly training their researchers and staff working with animals (3, 5).

In Europe, the first transnational regulation of animal experimentation was agreed in 1986. On March 18th, 26 countries agreed and signed in Strasbourg the European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes (ETS 123) (Council of Europe 1986). The convention legally bound the signatory member states and entered into force in 1st January 1991.

ETS 123 covered several areas of laboratory animal experimentation, including education and training, citing the term "competent" for personnel working with laboratory animals (6). The convention included two technical Appendices that were examined and revised in the subsequent years.

In November 1986, a more concise and objective document regarding the laboratory animal requirements - The Directive on the approximation of Laws, Regulations and Administrative Provisions of the Member States Regarding the Protection of Vertebrate Animals used for Experimental and other Scientific Purposes (86/609/EEC), emerged based on the previous convection. This was approved by the Council of Ministers of the European Community, and consequently compelled all members' states to transpose the established rules to their national legislation, with the extra adjustments considered necessary (7). The directive, similarly to the previous regulatory documents, refers to the need of appropriate education and training for persons who carry out experiments with animals, to ensure their scientific competence. The directive's article 7 establishes that only a competent person (or a person under direct responsibility of this person) should perform animal experimentation. Article 14, completes the provision by stating that "persons carrying out or supervising the conduct of experiments shall have received instruction in a scientific discipline relevant to the experimental work being undertaken and be capable of handling and taking care of laboratory animals; they shall also have satisfied the authority that they have attained a level of training sufficient for carrying out their tasks" (7).

The most relevant milestones with impact on Laboratory Animal Science training and education are represented on figure 1

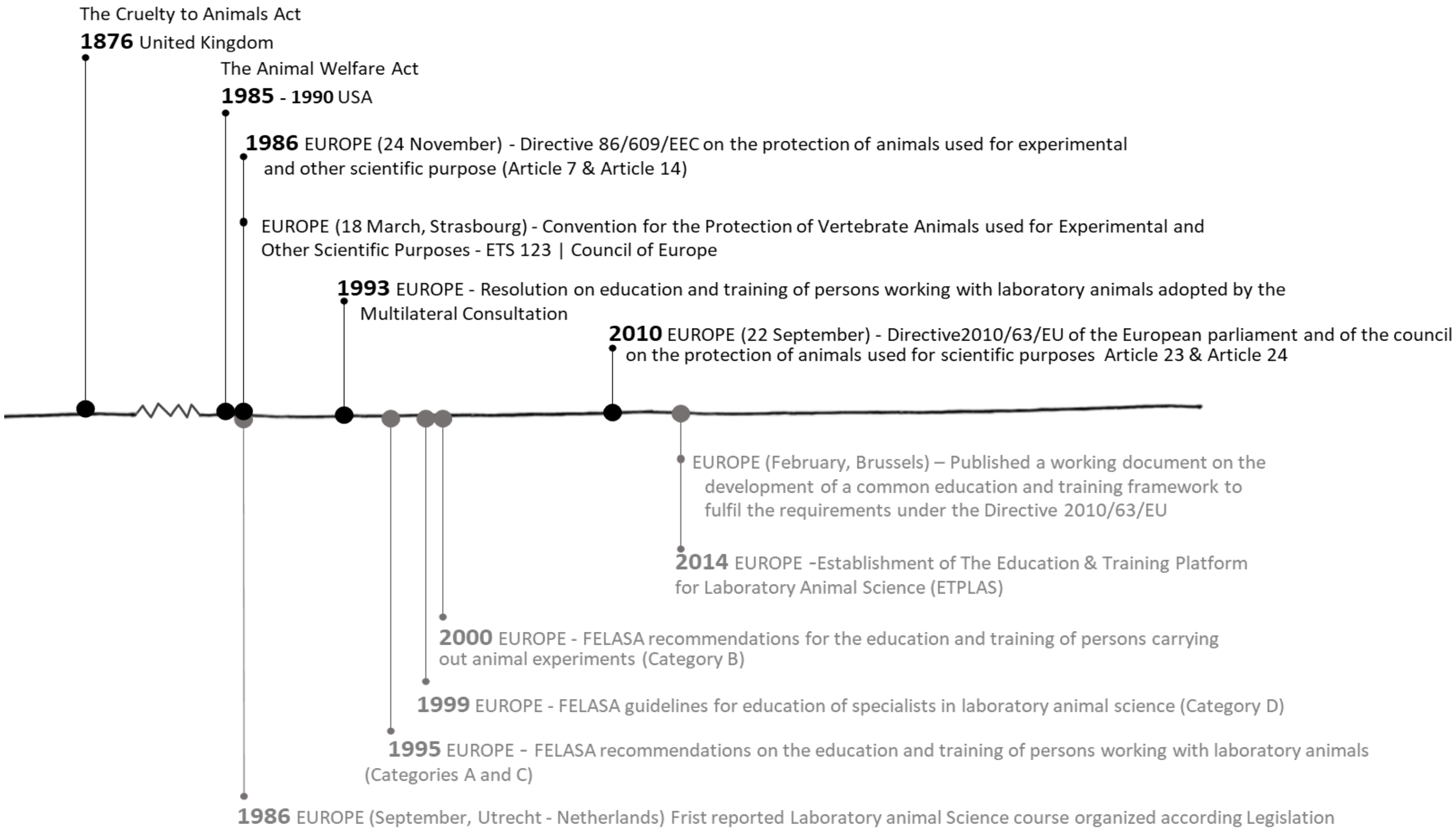

Figure 1 | Timeline summarizes the chronology indicating the relevant milestones with impact on Laboratory Animal Science training and education. Upper in the timeline is represented the main dates regarding the introduction of legislation and below the timeline are represented the emergence of recommendations and working documents relevant for the educational context.

The first laboratory animal science training course in Europe was reported in 1986. The origins of European training system started in Netherlands, which established that all graduate students planning to work with animals were required to participate in a course developed by scientists at Utrecht University (8). This course emerged with the legal requirement that made compulsory training for “new” researchers engaged in the design or performance of procedures with living animals. According to Rudacille, citing one of the course’ responsible, Professor L.F.M van Zutphen “the three-week course was developed in 1985 and organized for the first time in September 1986.” Course curriculum focused in ethical issues surrounding animal experimentation, the 3Rs – Replacement, Reduction and Refinement - putting emphasis on students’ attitude towards animals used in experimentation. The course was developed in strict collaboration between members of the department of veterinary faculty, with contributions from representatives of the government, scientific community, and animal welfare organizations. All the following courses had the same format and were coordinated centrally by the chair of laboratory animal science in Utrecht. In 1993, this course became international (8). The requirement of education in the laboratory animal science field would, during the following decades, give origin to the establishment of different courses and training programs at numerous European institutions and worldwide.

> “My personal feeling is that the course has a great impact on the scientist’ attitude towards animals. They are more sensitive to the issue of animal welfare and more willing to see possibilities to replace or reduce animal use, to seek methods and procedures that reduce suffering or to contribute to the welfare of animals”
>
> Professor L.F.M van Zutphen in The Scalpel and the Butterfly
> by Deborah Rudacille (2000)

Meanwhile, the European convention, signed in 1986, provided multilateral consultations after the introduction of ETS 123, necessary due to the progress of scientific knowledge and technical development (9). It is also relevant to refer that the Animal Procedures Committee, which advised the Home Office in the UK, published reports detailing recommendations for the education and training of persons involved in animal experiments (10). Also, the European Science Foundation (ESF) has claimed in 2001 that “Investigators and other personnel involved in the design and performance of animal-based experiments should be adequately educated and trained” (11).

2.1.2 | FELASA CONTRIBUTION FOR LABORATORY ANIMAL SCIENCE EDUCATION

FELASA - Federation of European Laboratory Animal Science Associations, was the first entity to establish guidelines for education and training in Europe (12, 13). This organization, established in 1978, is open to laboratory animal science organizations of Europe and beyond and currently includes at least 21 national or regional laboratory animal science associations, representing in at least 28 European countries (12, 13). Being the major association of laboratory animal science in Europe, FELASA has held observer status at the Council of Europe and at the European Union during the revision process of Appendix A (Guidelines for Accommodation and Care of Animals) to the European Convention (ETS 123) (6) in 1986 and followed in the development of the European Directives on the protection of animals used for scientific purposes (12). The association was also involved in the transposition and implementation of the new Directive with the establishment of expert working groups by the European Commission.

Based on the provisions established by the 86/609/EEC directive, FELASA started issuing, a couple of years later, recommendations for the education and training of the different categories of professionals, recognizing the need for a common and quality scheme of education and training that could frame and standardize these activities among all European members and professionals (12).

The relevance of a common training scheme had two main purposes, firstly to promote qualified education and training for professionals based on both minimum requirements and competence, and secondly, to allow the recognition of professionals' competence and facilitate the movement of personnel across different European members states (12, 14).

Between 1995 and 2000, FELASA issued recommendations for different categories of training for professionals working with laboratory animals:

- Category A (including progressive sublevels A0 to A2) for caretakers, persons taking care of animals (15, 16),
- Category B, for technicians, persons carrying out animal experiments (14),
- Category C for researchers, persons responsible for directing animal experiments (15) and,
- Category D for veterinaries, specialists in laboratory animal science (17).

With the introduction of the directive 2010/63/EU, the above FELASA categories became outdated and are being replaced. Nevertheless, the first FELASA recommendations and guidelines on education and training for A and C categories issued in 1995, were adopted and constituted the basis for the creation of LAS courses around Europe (18). Regarding former Category C, the recommendations were that candidates should hold a full university degree in a biomedical discipline and also underlined that the candidate competence should be assessed by examination or other form of competence verification (15). Former D category recommendations defined the competence of a specialist in laboratory animal science based on proper qualifications (according to proposed curricula) and experience acquired (17). In this particular document, for D category, the authors recognized that "the term competence, as it first appeared in the European guidelines, was rather poorly defined and hence resulted in some confusion", denoting already the existing difficulties in defining competence and how to achieve it with training programs (17). Similarly, the recommendations for former B category also referred the concept of acquired competence, highlighting that it should not be defined considering the duration of training and education, but the ability to carry out a task. Still, the working group suggested that 40 hours of planned training would normally be sufficient to establish a basic level of competence in a new candidate. More, this document clearly states the need of assessment by stating that "no award recognizing competence should be granted without a thorough evaluation of the candidate" (14).

The role of FELASA in the promotion of the high standards on education in laboratory animal science further includes a quality accreditation scheme of the LAS courses in Europe. The accreditation was introduced on 1st of January 2003, within the accreditation of the first courses, being this accreditation scheme recognised as the premier in Europe for LAS courses (19).

Indeed, in all previous recommendations FELASA endeavored to address the Competence issues on the context of training and clarify the designation of competent person made in Directive 86/609/EEC.

The Council of Europe addressed in 1993 in the resolution on Education and training adopted by the Multilateral Consultation parties to the convention ETS 123
the competencies required for personnel working with animals. The four categories proposed by FELASA were recognized in this resolution and published two years later (2, 15). Also, the term "competence" cited in FELASA recommendations, would be largely cited and recognized on official documents and in the revised Directive 63/EU on 2010. During the process of transposition and legal adoption of the requirements of European directive 2010/63 by different members states, FELASA continued to support and promote the application of new directive demands, to ensure courses quality and training harmonization (20).

2.1.3 | TRAINING AND COMPETENCE OF PERSONNEL ACCORDING 2010/63/EU DIRECTIVE FUNCTIONS

In 2002, the European Parliament called on the European Commission to revise the 1986 Directive for the protection of laboratory animals. This long process was concluded on 8th September 2010 and the new European Directive 2010/63 took effect on 1st January 2013 (21, 22). This directive introduces new aspects, particularly regarding training and education, in its Article 23 and 24, clarifying the guidelines to attain “competence” and be adequately educated and trained before performing animal procedures. The document identifies four categories of personnel and places clear responsibilities on different persons’ roles, functions A to D (4 different functions which did not match with the former FELASA categories):

(A) carrying out procedures on animals;
(B) designing procedures and projects;
(C) taking care of animals; or
(D) killing animals (22).

Thus, to support the requirements of the directive, in terms of training and education and competence assessment, all Member States and main stakeholder organisations were invited to nominate experts to develop an EU Education and Training Framework. In 2012, the Commission established an Expert Working Group (EWG) to prepare a document on the development of a common education and training framework to fulfil the requirements under the Directive. The aim was the development of common guidelines to standardize the transposition of the directive articles 23 and 24 and its requirements, facilitate the accomplishment of the requirements for competence of all those involved in use and care of animals for scientific purposes and free movement of personnel. Therefore, the expectations were that training programs should be flexible, available and accessible, affordable and of standardized quality. The generated guidelines complement the directive, by identifying output-driven quality standards for contents. They clarify and discuss topics such as supervision and competence assessment, as well as continuing professional development (CPD). Moreover, this document also refers general principles for the use of live animals in education and training. Learning should be continuous and lifelong, and persons working with animals are expected to be aware of new developments, techniques, and advances in the three Rs. This working document on the development of a common education and training framework is not compulsory and each Member State can decide whether and how it should be implemented. Guidance also indicates that institutions should have training programs, competence assessment, supervised training (divided in 5 sequential stages) and oversight of the delivery of training, as well as identify needs of additional training to ensure that staff are continuously trained and properly educated (23, 24).

Training is organized into modules: core modules which are key and essential modules, common to all (four) functions and additional modules, required only for specific functions. There are species-specific modules with contents adjusted to the species aim of the module. Each module has specific learning outcomes and lists knowledge and practical skills that trainee must acquire. Training is focused in competence development, which should be verified only after a period of direct supervision following initial task learning. Figure 2 summarizes the modules with the required topics for different functions (A-D).

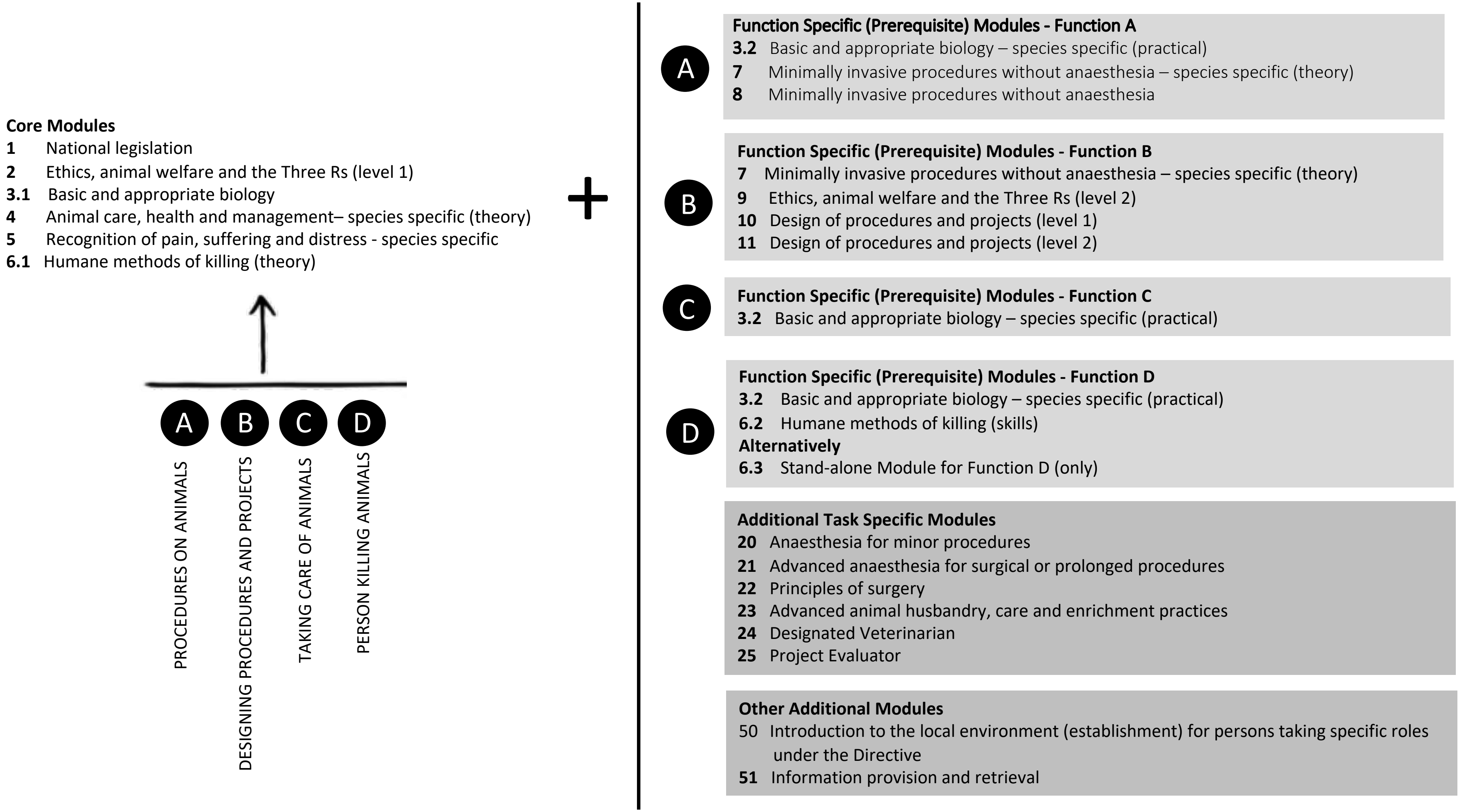

Figure 2 | The modular structure approach adopted by the Working Group (EWG) on Education and Training and the associated Learning Outcomes. The modules are grouped into three main categories: (1) *Core* represent the basic theoretical training for all personnel performing any of the Functions A-D of Article 23; (2) *Function Specific* (Prerequisite) required, in addition to the core modules to meet the minimum training needs for a specific function; (3) *Additional and Task Specific.*

In what Portugal is concerned, the Directive 2010/63 was transposed to Portuguese legislation by Decreto-lei 113/2013 on 7th August 2013 (25).

The EU Education and Training Framework recommended the creation of the Education and Training online Platform for Laboratory Animal Science (ETPLAS), which was established in 2014 to coordinate and develop a central and reliable education and training framework. With this information portal, with funding provided by the European Parliament in 2008, European Commission aims to provide tools for education of LAS community including course providers, assessors, staff or authority tasked with confirming validity of competence claims. The platform includes free and open resources such as e-learning modules based on the learning outcomes referred in the EU framework document; instruments to assess and verify competence, among others, some still under development (24, 26).

2.1.4 | UK ACCREDITATION TRAINING SYSTEM

The UK training accreditation system is different from the rest of Europe. Authorizations to perform animal experiments are regulated by the Animals in Scientific Procedures Act 1986 (ASPA 1986) (27) and associated legislation. Since 1994, Personal licenses (PIL: regulating the person), are required to the Home Office after the conclusion of modular training courses. Therefore, candidates to PIL need to undertake a mandatory training program, accredited by a recognized body according to a syllabus set by the Home Office. Currently, three accrediting bodies are recognized as Home Office for this accreditation purpose: the Royal Society of Biology Accreditation Board, the Scottish Accreditation Board or the Universities' Accreditation Group (28-31).

The courses curricula are structured around modules stated by the Home Office (Table 1), and the Home Office modular courses are now aligned to EU training modules. A PIL is obtained after modules completion and the types of experiments authorized are dependent of the modules completed. Candidates applying for a personal licence has to pass PIL A & B (EU Modules 1-8 & 20). The training process requires at least, 30 to 40 hours of study including lectures, practical training and written examinations. Modules are defined by the guidance on operation of ASPA 1986 and adapted by different institutions. Training certificates are valid for 5 years (32).

Table 1 | Brief list of topics explored in each module training required by the Home office in UK.

1	**LEGISLATION** Historical background, Ethical aspects of the use of animals, ASPA 1986
2	**PRINCIPLES OF ANIMAL CARE** Recognition of wellbeing, pain, suffering or distress, Euthanasia, Local procedures
3	**PRINCIPLES OF ANIMAL USE** Biology and husbandry, Diseases, Health monitoring, Anaesthesia and Analgesia, Minor procedures
4	**PRINCIPLES OF ANIMAL USE** Surgical Anesthesia and Analgesia, surgery
5	**PROJECT MANAGEMENT** Ethical considerations, Analysis of scientific literature, Alternatives and the 3Rs

Courses must include participation and lecturing by each institution' named veterinary surgeon' (NVS) and 'named animal care and welfare officer' (NACWO) which is normally a senior animal technician, and in the majority of the cases a member of the staff. At the end of training, the candidate is assessed and a formal certificate of modular training is issued. Feedback of the NVS, NACWO and other parties who have supported the candidate training is required to guarantee that the trainee has demonstrated the appropriate attitude towards animal use (28-31). The UK recognises FELASA accredited training, in that candidate only need to complete the modules that are relevant to national practice (32).

2.2 | EDUCATION IN LABORATORY ANIMAL SCIENCE: OVERVIEW AND APPROACHES

2.2.1 | OVERVIEW OF LABORATORY ANIMAL SCIENCE TRAINING

Effective education in laboratory animal science involves the acquisition of relevant knowledge and development of a suitable attitude. Training programs must be designed to allow: trainee acquisition of critical and technical expertise, development of a wide range of competencies according to the individual's role/category/function, and finally, to promote professionalism when working with animals. National and international animal research oversight bodies mandate that all persons involved in scientific procedures using animals, from the caretaker to the project responsible, must be supervised during the learning and training process until they have demonstrated the competence necessary for their functions. These competencies include: providing evidence of their knowledge, understanding the legal and ethical requirements for their role and being able to perform experiments minimizing the amount of trauma to the animals. Therefore, to promote professionalism, awareness for animal welfare, good science by the use of proper techniques, adjusted and contextualized educational approaches are necessary (2). These are principles aligned with the 3Rs (Replacement, Reduction, and Refinement), namely Refinement, since trained personnel will be better prepared to cope with animal issues such as welfare, pain, distress, housing among others (33).

Currently, a major part of the European Laboratory Animal Science (LAS) courses offer is directed for researchers and technicians, covering current directive functions A, B and D (former FELASA category C) and functions A and D (former FELASA category B) respectively. With the adoption of the directive (see Introduction 2.1.3), LAS training programs and courses contents were required to be species specific and FELASA website currently lists more than 30 accredited courses (covering from A to D functions) essentially in rodents (rats and mice), rabbits and zebrafish. Fish, cephalopods, dogs and pigs are also represented (34). The majority of courses were established under the previous FELASA categories system and moved into the functions system with the adoption of new directive in 2010[1] (34). Therefore, they essentially maintained the core structure design according to the guidelines FELASA issued between 1995 and 2000.

[1] Table indicating the total of courses accredited by FELASA under categories system that have moved into the Functions system and renewed the accreditation. Courses missing may "not moved" due to extinguishing or were merged to comply with the new requirements.

Data collected from FELASA webpage accessed at 28/12/2021 (https://felasa.eu/education-training/course-listings)

Giving the original FELASA recommendations, introductory courses, former FELASA category B, directed at technicians (covering essentially actual functions A and D) - should have approximately 40 hours duration, where half of the total training should involve practical work closely supervised by a competent person (14). Similarly, advanced courses, former FELASA category C, directed at researchers responsible for designing experiments with animals covering functions A, B and D, should have at least 80 hours or the equivalent, whether taken as a block or as modules, or the education and training accumulated by other acceptable means (15).

Laboratory animal science courses are often intensive and condensed in time, lasting from three to ten consecutive working days. The program design will be dependent on the target audience: researchers, technicians or animal caretakers. (18, 34).Typically, courses combine a theoretical component and practical sessions with living animals. The theoretical component format includes lectures, oral presentation, problem-solving assignments and case discussion while practical sessions include demonstrations and hand-on supervised training (35).

2.2.2 | LABORATORY ANIMAL SCIENCE TRAINING APPROACHES & LEARNERS

With the dissemination of the Laboratory Animal Science courses and the constant and emergent need of training to prepare more biomedical researchers for working with animals, innovative teaching formats are being adopted, especially to support the theoretical course components. Some approaches, as hybrid courses combining e-learning with practical sessions, are seen as a potential solution for the intensive and demanding LAS courses (36). A number of dimensions concur to the complexity of LAS education. One of the most challenging, is the heterogeneity of course participants enrolling the same training program, in terms of previous education, and professional experience with animal experimentation, posing challenges to the design of training and assessments. Also, in LAS, there are a diversity of practical techniques requiring technical ability from the trainee and there are several circumstances in which the same procedure can be performed in more than one correct way. Other particular feature is that courses aim to prepare trainees to work with different species and different strains of living animal models. Additionally, LAS

Designation	Courses accredited under the FELASA "Categories system"	Moved to Functions system
A + A0	2	0
B 5 Y	11	8
C 5 Y	17	11
D 10 Y	2	1
N.D.	1	0

tutors also face the same teaching challenges as other tutors in higher educational environments such as the divergent learning styles (e.g., kinesthetic, auditory, and visual), course language (foreign participants, english non-native learners), context (e.g., and hazardous environments), competence assessment, and generational experiences, for example.

Although there is an extensive literature on teaching methodologies, formats and media channels in higher education applied to biomedical courses, there is a lack of studies on how these formats and approaches can be adapted and whether they impact in LAS educational contexts. Kennedy and Kim Froeschl (2018) recognized that topics related with education in laboratory animal are receiving increasing attention in laboratory animal conferences (37). LAS training courses are being adjusted following the educational developments and adapted to the current circumstances, and to the demands and requirements worldwide, such as the creation of distance learning modules (36). Courses and programs' design and training adaptations must consider concepts of adult learning and strategies to promote their engagement. Learning can be either dependent, teacher centered (pedagogy) or independent, based on the learners' experience and intrinsic motivation (andragogy. Educational research has shown that a combination of both would be the best approach in higher educational contexts of medical education and veterinary (38-40). The same was suggested for laboratory animal training, keeping both approaches, leaving to the tutor the decision to adjust the method according to the demands (37). An adult learner's motivation is crucial to develop learning and motivation is strongly influenced by the educational environment (40, 41). Even in LAS courses, learner's motivation should be identified by the training coordination. In this context, the motivations can be so diverse as the need of license to work or the interest to learn a particular technique, for example. The use of that piece of information to structure training and define formats and approaches will certainly contribute to the training acceptance by the learners (37, 42). Consequently, trainee engagement during LAS training will be a key factor to overcome some of the LAS educational challenges (35).

A variety of educational methods and teaching formats are available and should be used and combined in order to match the course content, the trainee needs and the existing resources. The committee on educational programs in LAS (1991) cites that adult learners appreciate individualized, independent study, which can, for example, include videotapes, slides, and print visuals, also refering the use of audiovisual resources in this type of education (35, 43).

2.2.3 | THE USE OF ONLINE LEARNING FORMAT IN LAS CONTEXT

As consequence of the wide and fast development of internet access and multimedia technologies, e-learning has grown over the past two decades. It is now considered a real alternative to face-to-face classes in diverse situations, opening the range of options to deliver knowledge, namely in terms of channels, and timings (44-46). Nevertheless, to discuss the concept of online learning and properly contextualize the e-learning use in higher education, it is imperative to mention the impact of COVID-19 pandemic. The COVID-19 outbreak totally revolutionized the e-learning adoption, pressuring institutions to an unprecedented move of most of their services to online platforms. This speeded up the implementation of the online learning format, which was growing until 2020 approximately 15.4% per year in educational institutions around the world (47-51). Some authors even reasoned that emergency responses during crisis, such as the case during pandemic, should not be designated as online learning, but emergency remote education (ERE). Indeed, ERE might include the adaptation of traditional face-to-face classes to distanced learning (blended or fully) without making changes to the curriculum or the program. More, ERE, according to its definition, is suspended, with return to regular format once the crisis or emergency has been controlled (50, 52). Therefore, the type of learning adjustments established in response to the COVID-19 pandemic should not be considered proper online learning (50, 53).

The impact of adaptations made in higher education system as consequence of unprecedented disruption caused by the pandemic will require probably a few more years to be deeply assessed and understood. Nevertheless, a lot of what was previously established (pre-pandemic), regarding online learning was used during these months and this crisis enlightened the new online teaching opportunities, speeding up the implementation of distance approaches(50, 54).

Before the emergence of COVID-19 outbreak in the beginning of 2020, e-learning had already proved to be a useful tool and, when properly designed, capable of meeting the established educational goals. It has become part of programs in higher or post graduate education in medical, health or veterinary areas (44, 55, 56). A meta-analysis aiming to assess the impact of internet based learning in health professions education reported in 2008 the evidence of a significant positive effect across a variety of learning outcomes (57). Conceptually, e-learning was typically defined as the process of use of telecommunication technology (such as internet, intra- net/extranet, audio, video, satellite broadcast, inter- active TV, and CD-ROM) to extend learning or deliver instructional materials (58, 59). E-learning is also known as online learning, web-based learning, distributed learning, computer-assisted instruction or internet-based learning (44, 59).

Indeed, e-learning introduced several undeniable advantages when compared to traditional classroom learning. The most relevant are: the ability to be self-paced, flexible regarding time and location, providing unlimited access to knowledge to a potentially global audience and to be considered cost-effective for learners. The flexibility introduced by e-learning enabling students to balance their study, work, and social commitments is by far the most reported advantage. Nevertheless, e-learning presents specific disadvantages, such as lack of prompt feedback namely when asynchronous e-learning is adopted, increased tutor investment to prepare online materials and can be potentially frustrating, can increase anxiety and confusions for some users (46, 60).

Within the wide context of e-learning, particular terms and definitions have emerged with its growth to designate formats and combinations or designs using e-learning. The most used were Blended learning (BL), also known as hybrid learning to refer to a specific format which combines e-learning with traditional instructor-led classroom activities, resulting in a hybrid/mixed learning approach. Flipped learning (FL) also referred as flipped classroom, refers to a format where leaners firstly contact with the online material (outside the class), using the class time to prioritize active learning, by developing harder work of knowledge assimilation, through problem solving, discussion, debates, etc. (61, 62). It is also important to note that these concepts and definitions are also being adjusted, as they are being implemented and applied, even more in a "COVID era" (63, 64).

Yet, there are software applications or web-based technologies used in distance learning typically called a learning management system (LMS), course management system (CMS) or a virtual learning environment (VLE) allowing the share of educational materials and online interaction. Acronym VLE is used in Europe and LMS in North America (44). Moodle, Blackboard, WebCT and Totara are examples of popular learning management systems. Particularly, Moodle is a free and open-source used in laboratory animal science training at i3S.

Looking particularly at the impact of e-learning in the last decades in veterinary sciences – a research area to a large extent comparable to laboratory animal sciences- the number of studies published have been growing during the last decade. A survey conducted internationally to veterinary students in 2017 reported that students access a wide range of online veterinary resources to support their learning (65).

Several authors have investigated the acceptance and the students' feedback regarding the use of e-learning in veterinary sciences education. Particular findings regarding the learning characteristics and style of veterinary technology students enrolling distance-education program were identified, and according to the authors they should be considered when planning learning at distance. Results suggested that students enrolling in the distance learning program take more control of their own learning progress when compared with on-campus group. Thus, they preferred independent learning, participated more and were more motivated to learn, with problem solving and active experimentation being preferred by the most successful students. Also, the most independent and motivated students participated more and were more satisfied with learning, thus achieving better results (66). Although e-learning cannot fully replace particular teaching practices, it can help to attain learning outcomes which are difficult to reach with traditional teaching methods. An acceptance study with students and professionals of veterinary in German, considered e-learning in the example of the CASUS system an efficient teaching method (90% of the respondents) (67). Similarly, bachelor veterinary students in Czech Republic, positively evaluated (100% of respondents) the use of e-learning in "food safety and quality" program and recommended its use in other educational years (68). It was also reported a positive perception of online video teaching resources by Nottingham University veterinary students (69). A study comparing the experience of two cohorts of veterinary students reported that students under flipped classroom format achieved higher grades and satisfaction with learning resources, when compared with the traditional classroom students. No differences were described between cohorts regarding satisfaction with other aspects of the course (70). Learning outcomes were improved, as well as tutorial support efficacy when online tutoring support was introduced, in combination with face-to-face method in two basic subjects (cytology and histology; veterinary pharmacology), allowing a student-centered approach (71).

A growing body of research has also been reporting the impact of e-learning in the development of knowledge and skills at veterinary disciplines. Aiming to improve the transfer of knowledge, to offer contents that are more attractive and to allow unlimited access to the course, the Clinic for Ruminants of Veterinary Medicine at Vienna University created a multimedia e-learning course as additional tool to "Endoscopy in Ruminants" module. Study outcomes showed that the e-learning course brought several advantages particularly in the improvement of the teaching quality, in the promotion of the knowledge delivery and its adaptation to different learning types. Despite this e-learning tool being considered an ideal addition to clinical 'hands on', it was clear that it could not replace classical training and that this tool should only be considered as a course complement (72). An example which corroborates this fact was the creation of a teaching tool - the animated breathing pattern videotape, developed to support veterinary students to get sufficient clinical experience to observe all types of respiratory diseases and their resultant abnormal breathing patterns (73). To allow students to maximize their learning and preparation for Extramural clinical placement, a computer-aided learning (CAL) package titled ''The EMS Driving License'' was developed in the United Kingdom. Students recognized the e-learning training as relevant, able to address their concerns, and all of them recommended the training package (74). A qualitative study reported changes in study behavior and perception of learning, when e-learning activities were combined with traditional anatomy lectures. Students recognized that the online activities encouraged them towards a first engagement with the course, to be more active in class sessions, and eased their study for the assessment. They also considered that they would remember the course contents longer (75).

The use of Case-based e-learning (CBEL) as an alternative to traditional Case-based learning (CBL) applied in veterinary education, is another example of how e-learning is used to promote the development of clinical reasoning, critical thinking and problem-solving competencies. Second year veterinary students recognized that CBEL was effective and methodical approach to achieve the learning outcomes. Respondents also perceived CBEL as potentially promoting clinical confidence (55). More recently, e-learning was also able to mitigate the effects of 2020 lockdown measures due to COVID-19 pandemic, in the acquisition of compulsory "Day One Competences" achievement within the course "Inspection and control of food of animal origin" during the fourth year of the veterinary medicine degree program. Despite the authors identifying that practical training in this context could not be substituted by a fully online approach, the benefits of a combined strategy were recognized, since the support provided by the online materials was clear in the development of professional skills (76).

Until now, very few studies have explored the use e-learning and its influence in the development of practical skills in handling animals. Steinberg et al. (2007) introduced the development of a computer assisted learning (CAL), to act as a tool to deliver knowledge, through animation and simulation, addressing considerations of animal use to help to sensitize young scientists for humane handling of animals in research. The relevance of CAL development was supported by the documented need to harmonize specialist education and to create new ways of communicating important questions in laboratory animal sciences (77). A study to assess the impact of a teaching resource based on online videos to enhance knowledge of livestock handling, was conducted with veterinary and animal science students. The online audiovisual resource included a series of 68 animal handling videos with demonstrations of a variety of procedures, used in the handling of sheep, cattle, pigs, horses and poultry. Students acknowledged that the tool was a valuable complement for development of proficiency in safe and effective handling of animals. Results show that videos impacted students learning, by allowing the familiarization with the correct animals handling. The importance of a resource promoting a flexible learning in this context was also highlighted. Nevertheless, trainees also reported the need of more opportunities to practice animal handling using live animals (78).

E-learning is also been reported as a great resource to promote continuing professional development (CPD) and deliver knowledge to professionals in veterinary context. The need to maintain and stimulate expertise post–qualification is unquestionable, particularly in veterinary field, which is a context with quick changes and scientific advances concerning techniques and therapies. Traditional CPD programs raise issues such as location, time availability, cost, training structure, personal commitments, among others. Also, in this framework, the flexibility that e-leaning brings to the educational setting makes it a suitable channel to apply to CPD programs. In 2003, the Royal Veterinary College offered a series of e-CPD courses, online format exclusively, directed at veterinaries surgeons. During one year approximately, participants from 23 different countries reported satisfaction with the course format, claiming that e-learning can be an effective alternative to traditional CPD training (79).

Alessandrini et al. (2012) reported the e-learning use to emergency management when West Nile virus disease outbreaks were identified in northern Italy. Online training allowed the transfer of knowledge, skills and competence to approximately 450 veterinarians facing the epidemic. Online training was efficient in this scenario, to respond to an emergent need of veterinary training and control in emerging and re-emerging animals' diseases and zoonoses (80).

2.3 | COMPETENCE ASSESSMENT: HEALTH SCIENCES EDUCATION CONTEXT

2.3.1 | COMPETENCE DEFINITION

White (1959) reported one of the first definitions for the term "competence" as one's ability or capacity to interact effectively with its environment and carrying out goals (81).

Looking retrospectively, there were different definitions advanced for the competence concept, according authors and research fields. The implementation of this term into different contexts, activities and areas of knowledge - considering all the necessary term adjustments - revealed the complexity of the construct and reflected the nature of all the variables and factors affecting clinical competence (81-84). Research identified different competence construct dimensions (such as attitudes, values, charters attributes among others), which are not easily observed and measurable, as initially reported by the behavioral definitions advanced for the term (85).

Violato (2019) stated that competence in medicine and in other health professions, is composed of many and different variables, some of them measurable, others latent, and also factors relating behavior to environment (86). Fernandez and colleagues, reviewed the competence definitions provided for health sciences education and revealed that the construct settles on three principles: “(i) competence is an integration and combination of different types of resources that can be deployed into action; (ii) is closely associated with a specific set of problems or situations and (iii) its mobilization to solve a problem or execute a task within a specific context, is the decisive proof of its existence ” (85). In fact, the authors argued that a student deemed to be competent, must show adequate knowledge, skills and personal attitudes (85).

Enhancing the competence of laboratory animal professionals is a critical objective of hands-on training. The multiple definitions of the concept across areas of knowledge challenge the development of fit for purpose curricula and assessments. In laboratory animal science, establishing a consensus definition for competence would benefit the harmonization of training goals and certification standards across institutions. This thesis advances a working definition for professional competence in laboratory animal science, based on the definition advanced by Epstein and Hundert in 2002 for medical doctors. This is a consensual definition in medical education, which can be adjusted to laboratory animal science, since in both contexts’ standards for competence, delineate key technical, cognitive and emotional aspects of practice. Also, in both fields, competence definition must consider the existence of non-measurable and highly relevant factors.

2.3.2 | COMPETENCE DEVELOPMENT AND SKILL ACQUISITION

A model of skill acquisition, originally described in 1980, by Stuart Dreyfus, a mathematician and system analyst and Hubert Dreyfus, a philosopher, based on the study of chess players, air force pilots and army tanks drivers and commanders, has illuminated different research areas. The Dreyfus model established 5 levels of proficiency, a theoretical continuum of skill acquisition, describing learning as a process, in which learners move through distinct and sequential stages from novice, advanced beginner, competent, proficient and expert (the ultimate skilled stage)(87). Despite the long time elapsed since it was first reported, this model remains useful to comprehend individual skill acquisition process and define levels of ability. The Dreyfus model generalized into health education (88-90) particularly provided basis to understand the development of clinical nursing skills (91). Benner evaluated the Dreyfus model in nursery context with individuals in different career stages confirming the evidence reported by this model of skill acquisition (91, 92). The difference between the five stages of skill development, reflect essentially changes in two general aspects: the transition between abstract principles to the use of past, concrete experience, which allow the individual a wider range of responses and the perception of the task, previously seen as a sequence of formal operations to a whole task in which only certain parts are crucial (87, 91). The development of a skill requires time and practice, to reach the higher model stages – the most skilled ones, where individuals no longer require rules and guidelines to learn but also become prepared to deal with unexpected events (87). Indeed, skill acquisition models, emphasize the assumption that to successfully complete a task the individual will demand the combination of an integrated and undifferentiated whole, composed of external and internal resources, underpinning that the learner requires time and practice to acquire a skill and time to maturate the knowledge to be capable to act in a specific situation (85). Some authors have additionally proposed a model of skill acquisition based on adaptive learning approaches, assuming that individuals will learn and innovate in response to practical challenges, providing the best performance in highly demanding and inconstant workplaces, as the health care environment, overcoming what is called "Knowledge and skills gap" (93).

Competence across scenarios & context specificity

An individual's competence is situated within a specific scenario and context. Epstein described how "competence is contextual, reflecting the relationship between a person's abilities and the tasks he or she is required to perform in a particular situation in the real world" (84). Physicians' knowledge, skills and attitudes, are not tight or closed and cannot be assumed to be long-lasting to "serve the physician well" in all contexts (84). Indeed, the individual ability to perform a task can be affected not only when moving across different scenarios, from a simulated environment to an authentic scenario for example, but also vary inside the same context/situation due to variation of contextual factors (94, 95). Indeed, context specificity affects the performance of the same individual in different situations or with different problems, and the observation of a performance is poor predictor of others (96).Literature describes different approaches that can be used to comprehend the context specificity and how competence is applied from one case, or situation, to the next. Particularly in medical education, this is a recognized problem, and there is a significant body of literature aiming to clarify how trainees (residents) apply, for instance, basic knowledge, concepts and clinical reasoning to a new task or new settings (97-99). Context specificity reveals the importance of 'state' contrasting to "trait" theories of learning and behavior, and research suggesting that the individual "traits" will be context dependent. Thus, the situation, i.e. the state is a better predictor of behavior than the individual personality (96).

The literature reports two main strategies to facilitate transfer between contexts, often difficult for novices: (i) to decontextualize conceptual knowledge or basic science allowing its application to a new context and (ii) exposing learners to multiple contexts in which they can practice and apply the concepts (99). This last strategy is clearly the one applied to residency training in medical education and professional training in several other health sciences (100). Indeed, competence is transferable across sites, and increased exposure to different settings will contribute to consolidate accumulated competence and can effectively help to prepare individuals for independent practice (100-103). Studies have also reported that when moving into new contexts residents are challenged not only to adjust and optimize the acquired skills, but to develop new ones, sharpening self-regulation abilities and resilience development, which will be essential to achieve proficiency in professions of health areas (104, 105). More, transitions between different environments during learning (e.g. -transition within medical school from pre-clinical training to clinical training; transition from residency to the position of attending physician) are crucial to prepare future physicians to achieve autonomy (101, 102). Nevertheless, contextual changes may bring stress and negative impact in training (105).

A recent study, with residents moved under training programs between different clinical contexts, recognized the complexity of successful performances in different context. The capability of working and learning in a different scenario demands more than applying the acquired knowledge and the prior achieved skills to a new situation. A successful performance will also be dependent on how trainee will face the challenge of meeting their physiological and practical needs, followed by developing a sense of belonging and legitimacy. Moreover, trainees also need to develop the ability to recognize when the task is different, as well as integrating how and why those differences matter (100). Teunissen and collaborators also highlighted the role of tutors and trainers during training, referring the importance of their support in the development of trainee abilities in recognizing the contextual changes to promote competent performances in new settings (100).

The relation between competence and performance

The definition of competence and performance, and their inter-relation has been extensively explored, and a review of extensive literature revealed the confusion and unclear usage of the two terms (106).

Particularly in clinical contexts, the distinction between both terms, competence and performance, gained increased relevance since the incorrect application of the constructs impacts directly the quality of training and the reliability of examination systems, and consequently the assessment of professionals' performance (106, 107).

A literature review on medical competence showed that a relevant number of studies applied the terms, competence and performance, wrongly and reached invalid conclusions, highlighting this as a major problem for the quality assessment (107). This study, which explored the distinction between terms, raised the question of their relationship, suggesting two possible approaches outlined below in Table 2.

Table 2 | The relationship between competence and performance, represented the two distinct interpretations described in the literature. Table created and adapted from Rethans et al. (1990) (107).

COMPETENCE vs. PERFORMANCE

RELATED	**CONCEPTS ARE NOT SEPARATE BUT ARE SEPARATE MEASUREMENT POINTS OF THE SAME CONCEPT** RELATION | Competence considered limit of one's performance; Performance indicates competence and vice-versa; Performance can be observed whereas competence can only be inferred. Reasoning supporting the basis of all existing examination systems.
NOT RELATED	**TWO SEPARATE CONCEPTS** NO RELATION | In authentic scenarios doctors are influenced by motivational and situational variables. Authentic context is completely different from the examination situation.

Also, Newble (1992) explores the relation between the term competence and the term performance, in order to distinguish them. Competence is seen as the mastery of a body of crucial knowledge and the acquisition of relevant skills which includes components of 3 main domains; interpersonal, clinical and technical. Knowledge and skills are interconnected and useful for problem-solving. It is unquestionable the role of competence itself for the performance, but according to this model, illustrated in figure 3, the importance of attitudes influencing competence and performance cannot be underestimated (108).

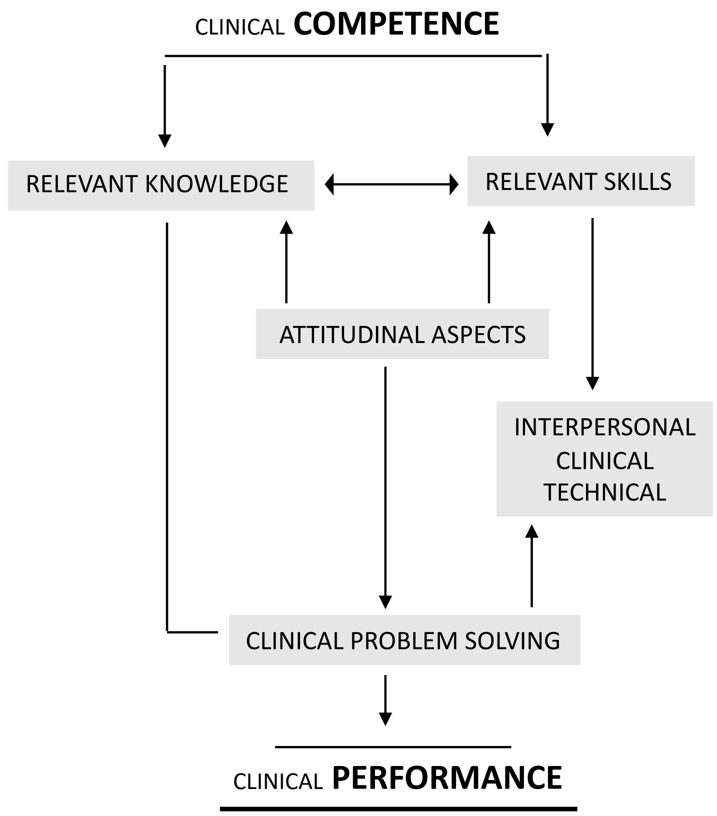

Figure 3 | Illustration of the components of clinical competence from Newble (1992) (108).

A great body of research defines competence in terms of what an individual should be able to do at an expected level or achievement, while performance is what someone actually does in real context. Therefore, competence would be what a person is capable of doing, and performance would be what is actually done in day-to-day practice (108-110). Some authors go further in proposing that competence is what a person knows and can do under ideal circumstances - that is potential, while performance is what is actually done in real life context - the actual contextualized behavior (111).

It is true that it is difficult to evaluate performance in practice (106, 112) as competence is not observable and therefore can only be measured indirectly. If this distinction can be less relevant while assessing competence in undergraduate scenarios, where the learner is not expected to work without supervisor, it can be critical in postgraduate areas in which performance assessment is required (108). Studies have evidenced that assessment of competence can be subjective and unreliable. In fact, the objectivity of a competence assessment in a high-stake test or in a clinical or similar workplace may be questioned, not only due to difficulties inherent to the context (scenario dynamism, expert judgement, etc.) but also because competence must meet standards conditioned and dependent on context. Benner (1984) asserted that “measurements can be only productive and accurate as the competences selected to be measured” (113-117).

It is empirically assumed that clinicians will be prepared and competent to work independently after finishing their residency or postgraduate training and also would remain competent throughout their professional life. Rethans et al. (2002) warn that this assumption does not reflect the reality, mainly for two reasons: (i) studies have shown that there are effective differences between what doctors do in controlled situations (during assessment) and daily practice (professional routine) and (ii) merely undertaking post-graduation won’t be enough to guarantee competence (118, 119). In fact, a growing body of research reported differences between the doctors’ competence demonstrated during assessment and behavior in daily professional routine (119-121).

In their review, Rethans and collaborators concluded that from the 18 studies analyzed, all, except one, fail to establish competence as a good predictor of performance, with researchers arguing that these results are justified with the poor quality of the measurement instruments (107). In contrast, a qualitative study that explored the relation between performance and competence of general practitioners showed that assessment of competence under examination circumstances can have predictive value for performance in actual practice if efficiency and consultation time are taken into account, besides that performance and competence are distinct constructs (122). Other research reported that the more the person’s competencies match the job requests, the more effective the person will be performing (123).

Kak et al. (2001) asserted that competence in health areas is the major determinant of performance; therefore, to measure competence, it is essential to ensure the ability and readiness of professionals. Nevertheless, performance will be affected by individual factors and external factors, with motivation assumed as a relevant aspect for a successful performance (124).

2.3.3 | COMPETENCE AND PERFORMANCE ASSESSMENTS

Trainee and student are regularly assessed in formal education with a variety of purposes (125). Assessments of professionals, particularly in clinical and veterinary areas, are also frequent. Competence assessment before entry into practice is considerably different from performance assessment in practice/authentic scenario (126). Typically, assessments are administered throughout a continuum of training and practice, since lack of practical competencies and knowledge may seriously compromise the expected performance (126). Looking at the assessments performed at higher education and during post-graduate training stages in these fields, they can be generally classified as formative or summative according to their purpose (127, 128). Formative assessments provide feedback, for both parties involved in the learning process – tutors and trainees, about the stage reached and what should be improved to achieve the expected level. The information provided by these assessments is crucial for learners' progress (127, 128). Summative assessments measure competence levels, generally at fixed timepoints (for example at the end of training periods) and support decisions about professional preparedness to practice or for independent work (127, 128). As an example, from the veterinary field, summative assessments are crucial to measure the so-called "day-one competencies". These describe the knowledge, skills and attributes required from veterinary students upon graduation, to ensure that they are prepared for their first professional activities, to be performed without any supervision (129, 130). However, it is important to highlight that, independently of the purpose, summative or formative, tests are hard to design (128).

Wass and collaborators (2001) stated that "assessment drives learning". Their main argument is that educational and formative assessments are of primary importance to what students focus their attention and efforts on (128).

Assessment Validity and Psychometric Properties

The quality of assessment methods or measurement instruments, depends on critical psychometric requirements: reliability, validity, feasibility or usability and credibility (86, 126). Hecker & Violato (2009) emphasized the importance of veterinary educators recognizing and understanding these concepts in order to promote assessment tools with evidence of utility, able to inform teaching practices, and able to guide curriculum development and revision (131).

For high-stakes decisions, assessments require strong reliability and validity while during training, educational programs tend to select feasible methods (less expensive regarding cost and labor) to judge (126). Reliability and validity are the most considered and investigated psychometric characteristics and are introduced below.

<u>Reliability</u>

Reliability is a measure of reproducibility or consistency of an instrument, and technically estimates the error measurement for assessment score (132). In other words, is the instrument ability to consistently score in different conditions of space and time. Reliability refers to how stable, consistent or accurate an instrument is (133, 134).
The reliability is influenced by factors as diverse as raters, assessment place/ stations used, candidate emotional state, sample characteristics and instrument administration conditions. Three aspects of reliability have been especially focused by research: the inter-rater reliability (consistency between examiners), intra-rater reliability (consistency of the same

examiner) and inter-case reliability (consistency of candidates across cases). Reliability is not a fixed property; therefore, the same instrument may not be reliable under different conditions (128, 134). A highly reliable instrument, of knowledge for example, with a standardized format will have a reliability of 0.90 or greater, a good score reproducibility, with a very low error rate. The reliability scores are expressed with values ranging between 0, unreliable and 1, perfectly reliable. In high-stakes decisions, based on performance and skills assessments, instruments must present a reliability above 0.85. More, in complex assessment contexts, such as simulations, which may combine different instruments, different scores of reliability must be estimated (126). A review directed at veterinary medical educators and administrators concluded that reliability supports the interpretation of assessment scores, and assessments demonstrating high reliability scores provide evidence that the scores are an accurate measurement of performance (135).

Validity

Validity reveals whether an assessment method is accurately measuring what it intends to measure, such as person's knowledge, skills abilities, etc. (126, 128, 136).

Validity does not mean that an indicated method is valid *per se*, but it refers to a specific context and sample. It is not an instrument characteristic and must be determined regarding a specific matter (133). Also, validity is a not a single statistical calculation but a property combining statistics, observations and arguments to support the quality of the validity evidence (126).

Validity is not totally independent from reliability, since some authors sustain that an instrument that is not reliable won't be valid. It may also occur that a reliable instrument (high reliable scores) is invalid (137).

Validity concept integrates main types of validity. Hecker & Violato (2009) consider four main levels of analysis when measuring the validity of assessments in Veterinary Education: 1) face validity, 2) content validity, 3) criterion-related validity (predictive or concurrent), and 4) construct validity (131).

Face and content validity can be established without empirical evidence. Face is considered the least sophisticated measure of validity, while criterion-related and construct validity require data for their demonstration, and therefore are more robust measures (131). The four types of validity are briefly described below.

I. Face validity has to do with appearance; does the instrument appear to measure what it is supposed to measure? It refers if the test appears (at face value) to measure what it claims to, from the respondent's view. This type of validity can be essential in establishing rapport and motivation, and in determining how seriously the assessment will be taken (131, 138).

II. Content validity refers to the degree to which the instrument content adequately reflects the construct/object that is being measured (126, 133). Hecker & Violato (2009) exemplify this validity with the following example: a client questionnaire that requesting his/her evaluation of the veterinarian's clinical skills and cognitive knowledge may fail and lack content validity, because these are not areas in which a client has sufficient expertise (131).

III. Criterion validity is applied when a performance on an assessment (or a score obtained with an assessment) correlates with a performance on some other external criterion. This criterion must be a widely accepted measure, with the same characteristics of the primary assessment tool, that is, an instrument or criterion considered 'gold standard' (131, 133, 134).

IV. Construct validity is the degree to which a group of variables really represents the construct to be measured. Abstract concepts are harder to qualify when establishing the construct validity. Some examples of measurement constructs in educational and health sciences are professionalism, communication skills, empathy, collegiality, and team participation. This validity requires the integration of information collected from several sources (131, 133).

Despite being less explored, feasibility or usability and credibility are also assessment psychometric requirements. Briefly, feasibility or usability is related most of the times with question budget, expertise, opportunity, resources, and time. Credibility is defined as the veracity of assessment, for the those who will use or benefit with the exam outputs (e.g. entities or organizations) (126).

Definitely, it is mandatory to guarantee that the quality of assessments that are developed and implemented are valid and reflect the examinee level of competence. Newble et al. (1994) proposed a set of guidelines for assessing clinical competence, highlighting four main sections that should be addressed with this purpose:

a) *Defining what is to be tested*. Considered the most important phase. Content validity (as well as other test properties) should be ensured at this point. The tasks which students are expected to be competent at this point should be identified and design the blueprint (the program plan with the skills expected of trainees)
b) *Selecting test methods*. Identify the most appropriate method to assess the task as well as understand the practical constrains of its administration.
c) *Addressing issues of test administration and scoring*

d) *Setting a standard for performance*. At this point the standard desired must be determined and consequently the procedure for result communications should be established (139).

Regarding the assessment of clinical competence, generally it can be divided in two types of scenarios: a ward or practice setting or an examination setting. Practice-based assessment is the most desirable environment to assess a student performance because it allows the possibility of registering multiple and direct observations, throughout a period of time in different situations. Nevertheless, and despite the advantage of these circumstances, in medical education it is rare to observe students as they perform patient histories and examinations. Therefore, even presenting high face validity, these assessments (with a small number of observations) are likely to be unreliable and unfair, and should only be used for formative purposes. In other hand, examination-based assessments, which includes a variety of methods of assessment presenting unsatisfactory levels of validity and/or reliability, depending on the method, has promoted the search of more complete examination alternatives, as the Objective Structured Clinical Examination. It is important to highlight that multiformat assessment is crucial in these clinical areas (108).

Assessment Methodology

The most common assessment methods used in health sciences and medical areas are listed on table 3. These methods can be grouped according to the focus of the measurement or to what parameter is intended to be measured with the assessment: knowledge, decision-making and judgment, performance and attitudes, skills and tasks.

Table 3 | The most common assessment methods used in health sciences and medical areas. Table created and adapted from Newble et al. (1992) and Bashook (2005) (108, 126).

METHODS OF ASSESSMENT

Knowledge

Assessing recall of facts, concepts, principles and basic application in a standard examination format.

Test formats:	Multiple Choice Questions
	Multiple Essay questions
	EQs
	Short answer questions

Decision making

Assessing the use of knowledge in realistic situations and decisions for critical actions.

Test formats:	Oral Exam (standardized) - e.g. Simulated live patients
	Key features case - e.g. written case focus on decisions

Performance and attitudes

Assessing skills requiring observing behaviour.

Test formats:	Global ratings
	Supervisor's narrative reports
	Feedback surveys
	Portfolios
	Standardized patient simulations (SPs)

Skills and tasks

Assessment of competencies involving specific skills (during training or work context)

	Global ratting scales and Checklists
	Standardized patients' simulations (SPs)
	Simulations and models
	Role-playing

Written tests containing items in diverse formats, such as MCQs, short-answers questions and essays, were traditionally the most used methods of summative assessment. They were considered important in earlier forms of assessment in health sciences (119) and are used to assess trainees' knowledge and problem-solving aspects of competence, by allowing to verify the recall of large amounts of facts (108, 127). MCQs have supplanted the use of long essays, due to easiness and objective correction (108). However, these methods, the written ones, present several limitations, and despite the fact that they fit the purpose to assess trainee knowledge, they revealed to be incomplete and unsuitable to assess trainee proficiency and skills progression. Oral examinations, used to assess decision making, also present psychometric limitations and were abandoned in the assessment of residents, being replaced by evaluation of clinical competence (the essential components, including clinical skills) through direct evaluation methods (140).

Looking retrospectively, in medical education more complex assessments strategies were adopted and developed gradually according to the confirmation of the insufficient response of the existing ones. Briefly, clinical evaluation exercise (CEX), was developed in collaboration between American Board of Internal Medicine (ABIM) and program directors, to create efficient and effective evaluation systems to be used in undergraduates and graduated medical education, but also revealed problems with reliability (140). An effort to combine ‘short case’ and ‘long case’ examinations were also used to assess performance and frequently, these methods also included an oral or viva examination (Sood 2001). Harden (1975) described the first objective structured clinical examination (OSCE). The OSCE emergence in the late 1970s and 1980s was perceived as a major revolution, by allowing the measurement of the students’ competence in an artificially created setting, with a design capturing more than previous instruments of assessment. OSCEs introduction was so relevant in the context and suitable for the purpose of the professional and clinical assessment that its introduction was followed by a period with no development of new instruments in health care areas (141).

In the 1990s, the discussion about reliability and validity of assessment instruments and their ability to measure competence and performance gained a new impulse. Indeed, experts highlighted the emergent need to implement assessment instruments capable of measuring not only (clinical) knowledge, but skills and professionalism (142, 143). Along these lines, instruments to assess competence, have been transferred in the last two decades from the traditional classroom environments to realistic settings such as clinical workplaces. Previous assessment instruments were being administrated to measure a trainee’s competence in artificially controlled environments, such the OSCE. Facing the need to perform the assessment of integral concepts and competencies at a realistic setting such as workplace, new instruments were developed (141, 144).

Workplace-based assessment refers to a group of assessment strategies which evaluate trainees’ performance at/during the clinical/real settings. There are a large number of workplace-based assessment instruments currently used in medicine, veterinary medicine and other health and clinical areas (141). The most reported are summed up at the table 4.

Table 4 | List of assessment instruments design to be used in the workplace learning context and assessment at undergraduate, graduate, post-graduate training levels.

WORKPLACE-BASED ASSESSMENT
Mini clinical evaluation exercise (miniCEX)
Direct observation of procedural skills (DOPs)
Multi-source feedback (MSF)
Case-based discussion (CBD)
Clinical encounters cards (CEC)
Clinical work samples (CWS)

Portfolios are also a useful tool, able to support assessment and competence development, however depending on the purpose are not accepted as assessment instruments (141, 145).

In the context of this thesis, and due to its potential applicability in laboratory animal science context, the most relevant assessment instruments reported for medical and health areas (145) are briefly enumerated below:

1) Objective structured clinical examination – OSCE,
2) Clinical evaluation exercise – miniCEX and
3) Direct observation of procedural skills – DOPS.

Objective Structured Clinical Examination – OSCE

The Objective Structured Clinical Examination (OSCE) in educational health science context measure student's competence in an artificially created environment with simulated scenarios/patients. It is not, therefore, a workplace-based assessment instrument. The major advantage of OSCE was the introduction of standardization and the control and reduction of the number of variables that could impact the assessment of performance. In a well-designed OSCE assessment, the trainees scores would be only affected by their performance, with minimal effect of external factors (146). To ensure the OSCE validity, the content should always be linked to the curriculum (147). Its structured nature, allowing to test different candidates under similar conditions, make students recognize the OSCE fairness (148, 149), being an exam with high face validity (150).

The OSCE was developed and first described by Harden et al. in 1975 as an objective and reliable alternative to the existing methods of assessing performance of medical candidates in clinically relevant activities (146, 147). Through its focus on skills rather than on knowledge, the OSCE overcomes the major weakness pointed at traditional evaluation models used at the time and improved the validity and reliability of assessments (147, 150, 151).

OSCE was successfully taken up in a wider range of health sciences including veterinary medicine; widely adopted for formative and summative assessments and certification purposes in most health professions, at different stages of undergraduate and post-graduate studies (150, 152-154).

In OSCEs, candidates rotate through a set of standardized stations in which they are observed while performing a clinical procedure. At each station, students are tested for a single competency, performing a specific task/procedure (147). Stations can eventually be inter-related or linked (150). The original format has been largely adapted with the time to the context and the requirements of assessments (146, 150, 151, 155). As examples, modified OSCEs such as the Royal College of Physicians' Practical Assessment of Clinical Examination Skills (PACES) and the objective structured long case (OSLER) were developed (147).

Stations might require performance supervision by one or more examiners – manned station or be stations where the candidate is unobserved - unmanned stations (150). At the manned stations, the examiner is equipped with measurement instrument as checklist or a global rating scale to rate the candidate performance - the reason of the objective descriptor in OSCE acronym (150). All candidates go through the same stations, and the exam is finished when candidates have completed all stations. An assessment blueprint ensures that stations reflect all aspects of the curriculum and learning outcomes (150). Generally, stations require three components: the "stem" (task) – which must be clearly stated, the checklist or the global rating scale, indicating the actions that should be taken by the candidate and the training information (147).

The concept of OSCEs for veterinary education is not new. According to Davis et al., one of the first studies that recommended it as an alternative to traditional examinations dates from 1993 (150, 154). It has been adopted in veterinary education at different stages and renamed as Objective Structured Practical Exam (OSPRE/OSPE), Objective Structured Practical Veterinary Exam (OSPVE) or Objective Structured Assessment of Technical Skills (OSATS) (89).

In the early years of training, veterinary OSCEs typically includes stations of animal handling skills, with different species. In the clinical years, other skills can be assessed such as placement of catheter, suturing, communication skills, etc. (89). Studies reported the introduction of the OSCE as a final-year assessment component in small animal internal medicine and surgery at the School of Veterinary Medicine (156, 157). In this approach, and after two successive experiences, this format proved to have acceptable reliability when compared with the previously used oral examination. Hence it was preferred, compared with an oral examination, by tutors and students (157). Hecker et al. (2010) developed an OSCE to assess first-year students' performance on procedural skills. Their preliminary results suggested that the OSCEs demonstrate face and content validity, and certain stations demonstrated adequate reliability, despite the fact that reliability issues were reported, related with first-time exam delivery, which might be improved (152). OSCE has been used in the summative assessment of the clinical competencies in ruminant health management of final-year DVM student at the Ontario Veterinary College (158). An OSCE to assess transrectal palpation skills for bovine pregnancy diagnosis, revealed to predict students' future pregnancy diagnosis accuracy (159). Also, the validated bovine transrectal palpation OSCE, supported an educational strategy to improve students' skills with the potential to reduce training time and animal use at teaching institutions (160). Previously, Lumbis et al. used OSCE to assess third-year veterinary students and compare the outcomes of two different learning approaches for periodontal treatment (161). Similarly, OSCE was used to assess proficiency of veterinary students, trained with an additional model developed to teach jugular vein technique injection in horses (162).

Nonetheless, a great part of these studies reported using OSCEs in veterinary focuses in the exam properties, namely the reliability, the validity and feasibility of this OSCE test format in the different veterinary contexts (127, 163-165), instead of exploring the advantages of this type of exam to the different veterinary areas and contexts. The OSCE exam reliability, tends to be related to the exam length and with the number of stations. Using many short stations, lasting approximately 1,5 to 2,5 hours generate scores that are sufficiently reliable for making pass-fail decisions than with fewer but longer exam stations (147, 150). Indeed, the high validity and reliability scores, of this type of exam, make it suitable for high-stakes summative examinations (150, 166). OSCE has also a positive educational impact (150).

Literature summarize the three main limitations pointed to this format. Firstly, the performance assessment of isolated aspects, due to the limited time per station. Secondly, OSCEs rely on task specific measurement instruments (checklists or Global Rating Scales), which may not be fully adapted to capture all performance. For example, a checklist might tend to emphasize thoroughness devaluing aspects of clinical experience. Also, there are limits to what can be simulated and it does not make it possible to assess performance in unpredictable scenarios (147).

[Mini] Clinical Evaluation Exercise - miniCEX

MiniCEX was originally developed in 1962, to assess clinical skills in medical training. The traditional CEX lasts about 2 hours and is conducted by a qualified physician who observes the trainee performance, assessing several aspects, such as the ability to collect a complete history and physical examination, in a pre-selected hospital patient. Trainees present their findings, regarding the diagnostic and therapeutic conclusions, in a written report, and receive directed feedback of the tutor afterwards (140, 167).

CEX presents three main limitations as instrument, which is independent of a resident's clinical competence. Firstly, trainee is observed by just one assessor which can bias the evaluation (168). Secondly, trainee observes only one clinical case, which may vary significantly and it is certainly one of the most relevant reasons for assessed performance not predicting further performances. The third argument lays on the fact that most of the physician-patient encounters are short and focused, while CEX, with its emphasis on all process could be a less relevant measure of clinical skill (167).

The reliability of the CEX is highly questioned in the literature, mainly due to its format in which each trainee performs a low number of CEXs – estimated one per year and is supervised by a single staff member (140). This limitations led ABIM to recommend the use of mini clinical evaluation exercise (miniCEX) (169).

The miniCEX is an adaptation of traditional CEX, contextualized with educational interactions and teaching rounds. The mini exercise is similar to the traditional one, however in the miniCEX the encounter is focused and lasts approximately 15 minutes. Several encounters are included in the overall assessment of a trainee, allowing assessment of performance with different patients and different contexts (i.e. ambulatory/out-patient, primary care). This design also allows the participation of multiple tutors in the assessment of the same trainee (167).

MiniCEX is very well accepted and has a positive educational impact reported by trainees and tutors (170, 171). Norcini et al. suggested that to produce a minimally reliable overall score, four to six Mini-CEX assessments were required (140). Other studies reported that acceptable reliability can be achieved with 10 encounters (172). Nevertheless, miniCEX is reported as a feasible and reliable evaluation tool (170, 173, 174). Particularly, it was identified a small component of variance due to trainee ability and a large component due to assessor variability – commonly pointed as the major limitation of this method. (170, 175, 176). Although miniCEX provides a global view of overall performance, the information regarding the measurement of single aspects are insufficient, to assess the overall candidate performance (177). This limitation turns miniCEX unfeasible for high-stakes /summative assessments (141). However, miniCEX has been adapted with success to the veterinary context (178).

Direct Observation of Procedural Skills - DOPS

The principle of direct observation of procedural skills (DOPS) is similar to miniCEX. It was developed in 2002 by the Royal College of Physicians in UK for workplace assessments of practical skills at graduate education (141, 179, 180). DOPS can be used to assess a wide variety of procedures. Typically, candidates have a determined time to perform the procedure under assessment on a real patient, directly observed by the rater, which gives feedback after performance. Effectively raters' timely feedback is seen as an advantage of these assessments by identifying the items that the candidate should improve. This prompt feedback can be a relevant education tool (181).

Instruments used to register DOPS may vary. Some institutions apply generic scores sheets (focusing in overall procedure aspects) while others design instruments to be specific for the procedure(179). Regarding psychometrics characteristics, research has shown that DOPS have reasonable reliability but require several observations to be a reliable measure of performance. Yet, fewer observations per year are necessary when compared with miniCEX, to obtain the same reliability (175).

Nevertheless, DOPS present some limitations, mostly related to implementation. Assessments require resources and time which might not be easy in clinical contexts. Observation time will vary according to the procedure assessed, but on average feedback time may take an additional third of the procedure observation time (179).The stress perceived by the candidate during assessment might be considered other limiting aspect (181). Regarding candidate's perception, DOPS are considered as a fair method of skills assessment, being practical and feasible (175). Currently, DOPS are widely used as assessment instrument in clinical and veterinary education for the assessment of competency in real-life situations (182). They are also implemented in undergraduate curricula of medicine (183) and veterinary (175) allowing the establishment of milestones for competence acquisition during training.

There are a few studies in veterinary context, exploring the use of DOPS. As example, a series of DOPS was used to assess a wide range of skills and attributes, including general clinical examination ability of 5^{th} year veterinary students at the University of Bristol. Results show that students perceived DOPS assessments to be beneficial to their learning (184). A qualitative study comparing the educational impact of DOPS and MCQs revealed that final year students of Veterinary Medicine and Science School at the University of Nottingham recognize that DOPS format has a positive influence on approaches to learning. Yet, students identify a conflict between preparing for final examinations and preparing for clinical practice (185). The majority of assessors which used DOPS to assess veterinary students' practical skills in final year clearly recognized the educational value of DOPS and would recommend it to another veterinary school. However, some have argued about the educational impact of DOPS and their ability to identify poor performances (182).

In Laboratory Animal Science, DOPS are being disseminated as a workplace-based assessment. LASA, the UK Laboratory Animal Science Association provides a list of DOPS in its website, intended to reflect current good practice. The association provides DOPS for laboratory animal science techniques, procedures and practical skills, and instructions to administer workplace-based assessments. The first tranche of these DOPS were created and shared with the association, by the Royal Veterinary College (186). DOPS for injection into mouse tail vein are also disseminated in its guide as a Competence Assessment Template, highlighting the importance of promoting objective and transparent ways of assessing performance in a practical task (187).

Miller's Pyramid Model of Assessment

George E. Miller proposed in 1990 a model for the assessment of clinical competence (188).The model known as the 'Miller pyramid' distinguished different hierarchical layers of competence therefore stating different conditions/contexts within which assessment should occur (188). With this new framework, Miller moved away from the traditional Flexnerian medical education model that was largely dominated by theoretical knowledge-based assessments, towards examinations based on clinical performance. The different layers of the Miller's model organize sequentially the stages of clinical competence and a distinction is made between de 4 different stages: *Knows* (knowledge), *Knows How* (competence), *Shows How* (performance) and *Does* (action) as shown in figure 4.

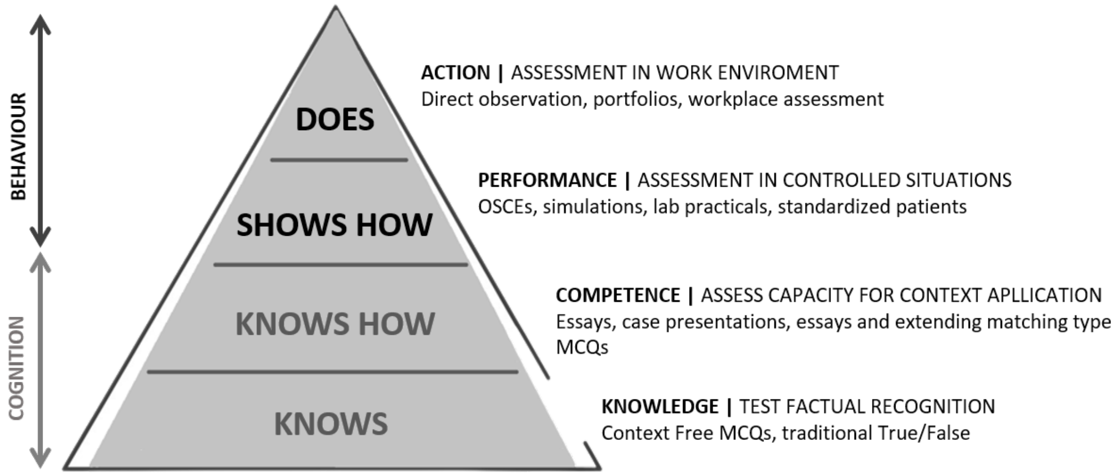

Figure 4 | Miller's Pyramid of Professional Competence with examples of assessment methods used in medical education, Miller's pyramid of clinical competence, modified from Miller (1990) (188).

According to Miller, it is reasonable to assume that either action or performance implies achievement of the more basic elements, knowledge and competence. The middle layers, competence and performance, are often used interchangeably. However, competence means that one can apply the knowledge to specific scenarios while performance is the ability to transfer the acquired knowledge and apply it to concrete actions. The top stages, show how and does, are therefore considered behavioral stages while the two initial stages are cognitive (179, 188).

Providing this assessment framework, Miller illustrates how skills can be measured, acknowledging, however, that no single assessment method is effective. Indeed, a fair judgement of so complex levels and behaviors as the one expected from physicians should not be supported by an assessment method only (188). The assessment of *Knows* recall the (factual) knowledge while *Knows how* demands the application of knowledge to problem-solving and decision making. These two first levels are assessed commonly by multiple choice questions, essays and oral evaluations. Therefore, in clinical areas in which trainees are expected to acquire proficiency, they would be appropriate mainly for early stages (128). The measurement of complex skills, at *Shows How* and *does* levels requires planned assessment formats to be informative (188). This framework, introducing the progress throughout the different stages and assessments methods must be largely transposed for veterinary medicine and therefore for laboratory animal science education.

Shows How and *Does*, the top stages of Millers pyramid, are the most challenging to measure and capture through assessments. *Shows How* stage can be assessed at controlled and simulated sets mainly through OSCEs, standardized patient-based tests and laboratory practices. The *Does* capture the ability in action and assessments at this stage will require performances in in vivo/authentic scenarios as direct observation, portfolios and workplace-based assessments (141, 189).

Standard Setting

"Livingston & Zieky (1982) define the standard setting as the process of deciding 'what is good enough?', 'how much is enough?" to set the standards which candidates must achieve during assessment (190, 191). The standard, also known as the minimum pass level, separates the competent trainees from those who are not. Standard setting is the process of determining this cut-score and is a fundamental step in assessment process (191-193). A standard setting should not be established arbitrarily and its main purpose is to remove the subjective factor (always present in case-by-case decisions making), standardizing the procedure to avoid that different decision-makers can attain different decisions considering the same basic information. The definition of the passing score implies prior definition of the standard for performance on the assessment, the minimum required for approval in that specific station (194).

Briefly, a large number of methods have been developed and used to set standards for both written and practical examinations (192, 193, 195). For example, Cusimano (1996) reported in his review more than 30 different criterion-reference standard setting methods, which he organized in 5 groups of methods: judgments of item characteristics, methods based on examinee performance, holistic impression of the examination, empirical statistical methods of adjusting standards and comprising methods (192).

Indeed, the process of defining a passing score (either using an absolute or relative standard) will, at some point, involve someone´s judgement. In order to guarantee that the decision is the most "correct" and the error is minor, these judgments must be: (1) made by persons with expertise, (2) meaningful to the evaluators and (3) made to guarantee that the purpose of the test is taken into account (196). So, in this sense, standard setting methods are dependent on the purpose of the test, and can be either norm-referenced (rank performances in relation to one another) or criterion-referenced (rank based on the number of correct answers provided by the student). Each of the methods serves a particular purpose and none is agreed upon as the best method or gold standard for all settings (194).

REFERENCES

1. Hajar R. Animal testing and medicine. Heart Views. 2011;12(1):42.
2. Taylor JF. Evolution of Laboratory Animal Program Management. In: Weichbrod RH, Thompson GAH, Norton JN, editors. Management of Animal Care and Use Programs in Research, Education, and Testing. 2nd edition. Boca Raton (FL): CRC Press/Taylor & Francis; 2018. Chapter 1. Available from: https://www.ncbi.nlm.nih.gov/books/NBK500428/ doi: 10.1201/9781315152189-1
3. de Greeve PCM, Hampson J, van Zutphen LFM. Legislation and animal experimentation. In: Zutphen LFM, Baumans V, Beynen AC, editors. Principles of Laboratory Animal Science, Revised Edition. The Netherlands: Elsevier Science B.V.; 2001.
4. International American Association for Accreditation of Laboratory Animal Care - History. Available from: https://www.aaalac.org/about/history/. (accessed 13 April 2022)
5. Adams B, Larson J. Legislative History of the Animal Welfare Act: National Agricultural Library - U.S. Department of Agriculture. 2014.Available from: https://www.nal.usda.gov/legacy/awic/legislative-history-animal-welfare-act-introduction.
6. Council of Europe. 1986. European Convention for the protection of vertebrate animals used for experimental and other scientific purposes (ETS123). Strasbourg: Council of Europe.
7. Council of the European Communities. 1986. Council Directive 86/609/EEC of 24 November 1986 on the approximation of laws, regulations and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes. Official Journal of the European Communities L 358/1.
8. Rudacille D. The Scalpel And The Butterfly, The Conflict Between Animal Research And Animal Protection: Arrangement with Farrar, Straus and Girpux; 2000.
9. Anderson D. Revision of Council of Europe Convention ETS 123 guidelines for the accommodation and care of animals used for experimental and other scientific purposes. Altern Lab Anim. 2004;32 Suppl 1A:183-5.
10. Report of the Animal Procedures Committee for 2006 - Review of Modular Training. London. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/231253/0041.pdf (2006, accessed 13 April 2022).

11. European Science Foundation. Use of Animals in Research . Available from: http://archives.esf.org/hosting-experts/scientific-review-groups/biomedical sciences-med/activities/science-policy/use-of-animals-in-research.html (2001, _accessed 13 April 2022)
12. Javier G. FELASA guidelines and recommendations. J Am Assoc Lab Anim Sci. 2012;51(3):11.
13. Prins JB. The History of FELASA – 1978-2018. Available online https://felasa.eu/Portals/0/FELASA_History_20180804-220406.pdf?ver=WVSsMJ41-D9. (accessed on April 13, 2022).
14. Nevalainen T, Dontas I, Forslid A, et al. FELASA recommendations for the education and training of persons carrying out animal experiments (Category B). Report of the Federation of European Laboratory Animal Science Associations Working Group on Education of Persons Carrying out Animal Experiments (Category B) accepted by the FELASA Board of Management. Laboratory animals. 2000;34(3):229-35.
15. Wilson MS, Berge E, Maess J, et al. FELASA recommendations on the education and training of persons working with laboratory animals: categories A and C. Reports of the Federation of European Laboratory Animal Science Associations Working Group on Education accepted by the FELASA Board of Management. Laboratory animals. 1995;29(2):121-31.
16. Weiss Convenor J, Bukelskiene V, Chambrier P, et al. FELASA recommendations for the education and training of laboratory animal technicians: category A: report of the Federation of European Laboratory Animal Science Associations Working Group on Education of Animal Technicians (Category A) accepted by the FELASA Board of Management. Lab Anim. 2010 Jul;44(3):163-9.
17. Nevalainen T, Berge E, Gallix P, et al. FELASA guidelines for education of specialists in laboratory animal science (Category D): Report of the Federation of Laboratory Animal Science Associations Working Group on Education of Specialists (Category D) accepted by the FELASA Board of Management. Laboratory animals. 1999; 33:1 - 15.
18. Carlsson HE, Hagelin J, Höglund AU, et al. Undergraduate and postgraduate students' responses to mandatory courses (FELASA category C) in laboratory animal science. Laboratory animals. 2001;35(2):188-93.
19. FELASA. Education & Training Accreditation Scheme. Available from: https://felasa.eu/education-training (2021, accessed 13 April 2022).

20. Gyger M, Berdoy M, Dontas I, et al. FELASA accreditation of education and training courses in laboratory animal science according to the Directive 2010/63/EU. Laboratory animals. 2019;53(2):137-47.
21. Hartung T. Comparative analysis of the revised Directive 2010/63/EU for the protection of laboratory animals with its predecessor 86/609/EEC - a t4 report. ALTEX. 2010;27(4):19.
22. European Union. 2010. Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. Official Journal of the European Union L276/33.
23. European Commission guidance document. National Competent Authorities for the implementation of Directive 2010/63/EU on the protection of animals used for scientific purposes. A working document on the development of a common education and training framework to fulfil the requirements under the Directive. Brussels, http://ec.europa.eu/environment/chemicals/lab_animals/pdf/guidance/education_training/en.pdf (2014, accessed 13 April 2020).
24. European Commission. Animals used for scientific purposes. Available from: https://ec.europa.eu/environment/chemicals/lab_animals/edu_training_en.htm (accessed 13 April 2022).
25. Decreto-Lei n.º 113/2013, Nº151 (2013).
26. Education and Training Platform for Laboratory Animal Science (ETPLAS). https://www.etplas.eu (2018, accessed 13 April 2022)
27. Animals (Scientific Procedures) Act 1986. London: HMSO.
28. Laboratory Animal Science Association. LASA position paper - Education and Training. Available from: https://www.lasa.co.uk/wp-content/uploads/2018/05/Position_Education-and-Training.pdf (accessed 12 December 2021).
29. Home Office. Guidance on the Operation of the Animals (Scientific Procedures) Act 1986. London: H.M.S.O, pp.3-48. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/662364/Guidance_on_the_Operation_of_ASPA.pdf. (accessed 12 December 2021).
30. Wright P. UK Regulations: How do you get a licence to carry out animal research? Speaking of Research. Available from: https://speakingofresearch.com/2013/03/25/uk-regulations-how-do-you-get-a-licence-to-carry-out-animal-research/ (accessed 13 April 2022).

31. Wolfensohn S, Lloyd M. Education and Training for the Personal Licence Holder. In: Wiley-Blackwell, editor. Handbook of Laboratory Animal Management and Welfare. 4 ed 2013.
32. Medical Science Divison University of Oxford. Personal Licence Training (PIL A, B & C).Available from: https://www.medsci.ox.ac.uk/study/biomedical-training-courses/pil (accessed 13 April 2022)
33. NC3Rs (National Centre for the Replacement Refinement & Reduction of Animals in research). The 3Rs. Available from: https://nc3rs.org.uk/the-3rs (accessed 13 April 2022)
34. FELASA. Courses listing - accredited under the FELASA. Available from: https://felasa.eu/education-training/course-listings (2021, accessed 13 April 2022).
35. National Research Council (US) Committee on Educational Programs in Laboratory Animal Science. How to Plan and Implement a Training Course. Education and Training in the Care and Use of Laboratory Animals: A Guide for Developing Institutional Programs. Washington (DC): National Academies Press (US); 1991.
36. Gyger M. Impact of the coronavirus disease 2019 pandemic on education and training: A FELASA first experience. Laboratory animals. 2020;54(4):397-8.
37. Weichbrod RH, Thompson GA, Norton JN, eds. Management of Animal Care and Use Programs in Research, Education, and Testing. 2nd ed. Boca Raton (FL): CRC Press/Taylor & Francis; 2018. p. 221-66.
38. Parkinson TJ, St George AM. Are the concepts of andragogy and pedagogy relevant to veterinary undergraduate teaching? J Vet Med Educ. 2003;30(3):247-53.
39. Misch DA. Andragogy and medical education: are medical students internally motivated to learn? Advances in health sciences education : theory and practice. 2002;7(2):153-60.
40. Taylor DCM, Hamdy H. Adult learning theories: implications for learning and teaching in medical education: AMEE Guide. Medical teacher. 2013;35(11):e1561–e72.
41. Sobral DT. What kind of motivation drives medical students' learning quests? Medical education. 2004;38(9):950-7.
42. Mikkonen J, Ruohoniemi M. How do veterinary students' motivation and study practices relate to academic success? J Vet Med Educ. 2011;38(3):298-304.
43. National Research Council (US) Committee on Educational Programs in Laboratory Animal Science. How to Approach the Task of Education. Education and Training in the Care and Use of Laboratory Animals: A Guide for Developing Institutional Programs. Washington (DC): National Academies Press (US); 1991.
44. Ellaway R, Masters K. AMEE Guide 32: e-Learning in medical education Part 1: Learning, teaching and assessment. Med Teach. 2008;30(5):455-73.

45. Kenzig MJ. Lost in Translation: Adapting a Face-to-Face Course Into an Online Learning Experience. Health Promotion Practice. 2015.
46. Zhang D, Zhao JL, Zhou L, et al. Can e-learning replace classroom learning? Commun ACM. 2004;47(5):75-9.
47. Alqahtani AY, Rajkhan AA. E-Learning Critical Success Factors during the COVID-19 Pandemic: A Comprehensive Analysis of E-Learning Managerial Perspectives. Education Sciences. 2020;10(9):216.
48. Stub ST. Countries Face an Online Education Learning Curve. USNEWS Available from: https://www.usnews.com/news/best-countries/articles/2020-04-02/coronavirus-pandemic-tests-countries-abilities-to-create-effective-online-education (2020, accessed 12 February 2022).
49. Bashir A, Bashir S, Rana K, et al. Post-COVID-19 Adaptations; the Shifts Towards Online Learning, Hybrid Course Delivery and the Implications for Biosciences Courses in the Higher Education Setting. Frontiers in Education. 2021;6.
50. Farnell, T., Skledar Matijević, A., Šćukanec Schmidt, N. The impact of COVID-19 on higher education: a review of emerging evidence NESET report, Luxembourg: Publications Office of the European Union. 2021. doi: 10.2766/069216.
51. Zalat MM, Hamed MS, Bolbol SA. The experiences, challenges, and acceptance of e-learning as a tool for teaching during the COVID-19 pandemic among university medical staff. PloS one. 2021;16(3):e0248758.
52. Shin M, Hickey K. Needs a little TLC: examining college students' emergency remote teaching and learning experiences during COVID-19. Journal of Further and Higher Education. 2021;45(7):973-86.
53. Hodges C, Moore S, Lockee B, et al. The Difference Between Emergency Remote Teaching and Online Learning. 2020. Available from: https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning (accessed 12 April 2022)
54. Gaebel M, Zhang T, Stoeber H, et al. Digitally enhanced learning and teaching in European higher education institutions. European University Association . Available from: https://eua.eu/resources/publications/954:digitally-enhanced-learning-and-teaching-in-european-higher-education-institutions.html (2021, accessed 12 April 2022)
55. Sawras M, Khosa D, Lissemore K, et al. Case-Based e-Learning Experiences of Second-Year Veterinary Students in a Clinical Medicine Course at the Ontario Veterinary College. J Vet Med Educ. 2020;47(6):678-94.

56. Regmi K, Jones L. A systematic review of the factors – enablers and barriers – affecting e-learning in health sciences education. BMC Medical Education. 2020;20(1):91.
57. Cook DA, Levinson AJ, Garside S, et al. Internet-based learning in the health professions: a meta-analysis. Jama. 2008;300(10):1181-96.
58. Holsapple CW, Lee-Post A. Defining, Assessing, and Promoting E-Learning Success: An Information Systems Perspective. Decision Sciences Journal of Innovative Education. 2006;4(1):67-85.
59. Ruiz JG, Mintzer MJ, Leipzig RM. The impact of E-learning in medical education. Academic medicine : journal of the Association of American Medical Colleges. 2006;81(3):207-12.
60. Cook DA. Web-based learning: pros, cons and controversies. Clin Med (Lond). 2007;7(1):37-42.
61. Vallée A, Blacher J, Cariou A, et al. Blended Learning Compared to Traditional Learning in Medical Education: Systematic Review and Meta-Analysis. J Med Internet Res. 2020;22(8):e16504.
62. Berrett D. How 'Flipping' the Classroom Can Improve the Traditional Lecture. The Chronicle of Higher Education [Internet]. 2012; 31:[A16-8 pp.].
63. National Education Association. Overview: Hybrid Learning Models .USA Available from: https://www.nea.org/professional-excellence/student-engagement/tools-tips/overview-hybrid-learning-models (2021, accessed 12 April 2022)
64. Nathaniel TI, Black AC. An Adaptive Blended Learning Approach in the Implementation of a Medical Neuroscience Laboratory Activities. Medical Science Educator. 2021;31(2):733-43.
65. Gledhill L, Dale VHM, Powney S, et al. An International Survey of Veterinary Students to Assess Their Use of Online Learning Resources. J Vet Med Educ. 2017;44(4):692-703.
66. Varnhagen CK, Wright DL. Learning characteristics of veterinary technology students in a distance-education and an on-campus program. J Vet Med Educ. 2008;35(3):449-55.
67. Borchers M, Tipold A, Pfarrer C, et al. Acceptance of case-based, interactive e-learning in veterinary medicine on the example of the CASUS system. Tierarztl Prax Ausg K Kleintiere Heimtiere. 2010;38(6):379-88.
68. Bartova E, Halova D. E-Learning for Veterinary students Houska M, Krejci I, Flegl M, editors. Prague 6: Czech University Life Sciences Prague; 2014. 27-31 p.
69. Roshier AL, Foster N, Jones MA. Veterinary students' usage and perception of video teaching resources. BMC Medical Education. 2011;11(1):1.

70. Dooley LM, Frankland S, Boller E, et al Implementing the Flipped Classroom in a Veterinary Pre-clinical Science Course: Student Engagement, Performance, and Satisfaction. J Vet Med Educ. 2018;45(2):195-203.
71. García-Iglesias MJ, Pérez-Martínez C, Gutiérrez-Martín CB, et al. Mixed-method tutoring support improves learning outcomes of veterinary students in basic subjects. BMC Veterinary Research. 2018;14(1):35.
72. Bernkopf M, Franz S, Baumgartner W. Experiences with a blended learning course for clinical veterinary education at the University of Veterinary Medicine Vienna, Austria. Tieraerztl Prax Ausg Grosstiere Nutztiere. 2010;38(2):99-108.
73. Hawkins EC, Hansen B, Bunch BL. Use of animation-enhanced video clips for teaching abnormal breathing patterns. J Vet Med Educ. 2003;30(1):73-7.
74. Bell C, Baillie S, Kinnison T, et al. Preparing Veterinary Students for Extramural Clinical Placement Training: Issues Identified and a Possible Solution. J Vet Med Educ. 2010;37(2):190-7.
75. Lochner L, Wieser H, Waldboth S, et al. Combining traditional anatomy lectures with e-learning activities: how do students perceive their learning experience? Int J Med Educ. 2016;7:69-74.
76. Tinacci L, Guardone L, Giusti A, et al. Distance Education for Supporting “Day One Competences” in Meat Inspection: An E-Learning Platform for the Compulsory Practical Training of Veterinarians. Education Sciences. 2022;12(1):24.
77. Steinberg N, von Keutz E, Weiss J, et al. E-Learning – A new tool for the education of young scientists in the humane treatment of experimental animals: A contribution to the 3R. 6th World Congress on Alternatives & Animal Use in the Life Sciences; Tokyo, Japan: AATEX; 2007. p. 285-9.
78. Klupiec C, Pope S, Taylor R, et al. Development and evaluation of online video teaching resources to enhance student knowledge of livestock handling. Australian Veterinary Journal. 2014;92(7):235-9.
79. Short N, Maddison J, Mantis P, et al. Veterinary e-CPD: a new model for providing online continuing professional development for the veterinary profession. J Vet Med Educ. 2007;34(5):689-94.
80. Alessandrini B, D'Albenzio S, Turrini M, et al. Emergency management: e-learning as an immediate response to veterinary training needs. Vet Ital. 2012;48(2):219-25.
81. White RW. Motivation reconsidered: the concept of competence. Psychol Rev. 1959;66(5):297-333.
82. Hubbard JP, Levit EJ, Schumacher CF, et al. An objective evaluation of clinical competence: new technics used by the national board of medical examiners. N Engl J Med. 1965;272:1321-8.

83. Clinical competence in internal medicine. Ann Intern Med. 1979;90(3):402-11.
84. Epstein RM, Hundert EM. Defining and assessing professional competence. Jama. 2002;287(2):226-35.
85. Fernandez N, Dory V, Ste-Marie LG, et al. Varying conceptions of competence: an analysis of how health sciences educators define competence. Medical education. 2012;46(4):357-65.
86. Violato C. Assessing Competence in Medicine and Other Health Professions 2018. CRC Press.
87. Dreyfus SE, Dreyfus H. A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition. Distribution. 1980:22.
88. Slater RJ, Castanelli DJ, Barrington MJ. Learning and teaching motor skills in regional anesthesia: a different perspective. Reg Anesth Pain Med. 2014;39(3):230-9.
89. Baillie S.; Booth N, Catterall A, et al. A Guide to Veterinary Clinical Skills Laboratories 2015. Available from: http://uilis.unsyiah.ac.id/oer/files/original/6b3baac9932416a16e3f90a03d43b709.pdf (accessed 12 April 2022)
90. Lyon LJ, Hoover TE, Giusti L, et al. Teaching Skill Acquisition and Development in Dental Education. J Dent Educ. 2016;80(8):983-93.
91. Benner P. From novice to expert. Am J Nurs. 1982;82(3):402-7.
92. Benner P. Using the Dreyfus Model of Skill Acquisition to Describe and Interpret Skill Acquisition and Clinical Judgment in Nursing Practice and Education. Bulletin of Science, Technology & Society. 2004;24(3):188-99.
93. Cutrer WB, Miller B, Pusic MV, et al. Fostering the Development of Master Adaptive Learners: A Conceptual Model to Guide Skill Acquisition in Medical Education. Academic medicine : journal of the Association of American Medical Colleges. 2017;92(1):70-5.
94. Ratcliffe T, McBee E, Schuwirth L, et al. Exploring Implications of Context Specificity and Cognitive Load in Residents. MedEdPublish. 2017;6(48).
95. Van Der Vleuten CP. The assessment of professional competence: Developments, research and practical implications. Advances in health sciences education: theory and practice. 1996;1(1):41-67.
96. Eva KW. On the generality of specificity. Medical education. 2003;37(7):587-8.
97. Dyre L, Tolsgaard MG. The gap in transfer research. Medical education. 2018;52(6):580-2.
98. Durning SJ, Artino AR, Boulet JR, et al.The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning

performance in experts?). Advances in health sciences education : theory and practice. 2012;17(1):65-79.
99. Kulasegaram KM, Chaudhary Z, Woods N, et al. Contexts, concepts and cognition: principles for the transfer of basic science knowledge. Medical education. 2017;51(2):184-95.
100. Teunissen PW, Watling CJ, Schrewe B, et al. Contextual competence: how residents develop competent performance in new settings. Med Educ. 2021; 55(9): 1100- 1109.
101. Atherley A, Dolmans D, Hu W, et al. Beyond the struggles: a scoping review on the transition to undergraduate clinical training. Medical education. 2019;53(6):559-70.
102. Westerman M, Teunissen PW, van der Vleuten CPM, et al. Understanding the transition from resident to attending physician: a transdisciplinary, qualitative study. Academic medicine : journal of the Association of American Medical Colleges. 2010;85(12):1914-9.
103. Lejonqvist G-B, Eriksson K, Meretoja R. Evidence of clinical competence. Scandinavian Journal of Caring Sciences. 2012;26(2):340-8.
104. Gordon L, Teunissen PW, Jindal-Snape D, et al. An international study of trainee-trained transitions: Introducing the transition-to-trained-doctor (T3D) model. Medical Teacher. 2020;42(6):679-88.
105. Bernabeo EC, Holtman MC, Ginsburg S, et al. Lost in transition: the experience and impact of frequent changes in the inpatient learning environment. Academic medicine : journal of the Association of American Medical Colleges. 2011;86(5):591-8.
106. While AE. Competence versus performance: which is more important? Journal of advanced nursing. 1994;20(3):525-31.
107. Rethans JJ, Van Leeuwen Y, Drop R, et al. Competence and Performance: Two Different Concepts in the Assessment of Quality of Medical Care. Family Practice. 1990;7(3):168-74.
108. Newble DI. Assessing clinical competence at the undergraduate level. Medical education. 1992;26(6):503-11.
109. SeniorJR.. Toward the measurement of competence in medicine. Philadelphia: National Board of Medical Examiners. Stern, D. T. (Ed.). 2006
110. Neufeld V, Norman GR. Assessing clinical competence. New York: Springer Pub. Co.; 1985.
111. Messick S. The Psychology of Educational Measurement. Journal of Educational Measurement. 1984;21(3):215-37.
112. Girot EA. Assessment of competence in clinical practice-a review of the literature. Nurse Educ Today. 1993;13(2):83-90.

113. Salvatori P. Clinical Competence: A Review of the Health Care Literature with a Focus on Occupational Therapy. Canadian Journal of Occupational Therapy. 1996;63(4):260-71.
114. Benner P. From novice to expert: excellence and power in clinical nursing practice. AJN The American Journal of Nursing. 1984;84(12).
115. Black H. and Wolf, A. Knowledge and Competence: Current Issues in Training and. Education, HMSO (Careers and Occupational Information Centre); London.1990.
116. Van der Vleuten CP, Norman GR, De Graaff E. Pitfalls in the pursuit of objectivity: issues of reliability. Medical education. 1991;25(2):110-8.
117. Ten Cate O, Regehr G. The Power of Subjectivity in the Assessment of Medical Trainees. Academic medicine: journal of the Association of American Medical Colleges. 2019;94(3):333-7.
118. Davis DA, Thomson MA, Oxman AD, et al. Changing physician performance. A systematic review of the effect of continuing medical education strategies. Jama. 1995;274(9):700-5.
119. Rethans JJ, Norcini JJ, Barón-Maldonado M, et al. The relationship between competence and performance: implications for assessing practice performance. Medical education. 2002;36(10):901-9.
120. Hamdy H, Prasad K, Anderson MB, et al. BEME systematic review: predictive values of measurements obtained in medical schools and future performance in medical practice. Med Teach. 2006;28(2):103-16.
121. Ram P, van der Vleuten C, Rethans JJ, et al. Assessment in general practice: the predictive value of written-knowledge tests and a multiple-station examination for actual medical performance in daily practice. Medical education. 1999;33(3):197-203.
122. Rethans JJ, Sturmans F, Drop R, et al. Does competence of general practitioners predict their performance? Comparison between examination setting and actual practice. BMJ (Clinical research ed). 1991;303(6814):1377-80.
123. Spencer LM, McClelland DC, Spencer SM. Competency assessment methods. University of Michigan: Hay/McBer Research Press; 1994.
124. Kak N, Burkhalter B, Cooper MA. Measuring the Competence of Healthcare Providers. QA Operations Research Issue Paper. 2001;2(1):23.
125. Earl L. Assessment as Learning: Using Classroom Assessment to Maximize Student Learning. 2nd ed. Thousand Oaks, CA2013.
126. Bashook PG. Best practices for assessing competence and performance of the behavioral health workforce. Adm Policy Ment Health. 2005;32(5-6):563-92.
127. May SA, Head SD. Assessment of technical skills: best practices. J Vet Med Educ. 2010;37(3):258-65.

128. Wass V, Van der Vleuten C, Shatzer J, et al. Assessment of clinical competence. The Lancet. 2001;357(9260):945-9.
129. Welsh PJK, Jones LM, May SA, et al. Approaches to defining day-one competency: a framework for learning veterinary skills. Rev Sci Tech. 2009;28(2):771-7.
130. Royal College of Veterinary Surgeons. Day One Competences. [16 p.]. Available from: https://animalowners.rcvs.org.uk/document-library/day-one-competences/ (2020, accessed 28 february 2022)
131. Hecker K, Violato C. Validity, reliability, and defensibility of assessments in veterinary education. J Vet Med Educ. 2009;36(3):271-5.
132. Brennan RL. An Essay on the History and Future of Reliability from the Perspective of Replications. Journal of Educational Measurement. 2001;38(4):295-317.
133. Souza AC, Alexandre NMC, Guirardello EB. Psychometric properties in instruments evaluation of reliability and validity. Epidemiol Serv Saude. 2017;26(3):649-59.
134. Keszei AP, Novak M, Streiner DL. Introduction to health measurement scales. Journal of Psychosomatic Research. 2010;68(4):319-23.
135. Royal KD, Hecker KG. Understanding Reliability: A Review for Veterinary Educators. J Vet Med Educ. 2016;43(1):1-4.
136. Roberts P, Priest H. Reliability and validity in research. Nursing Standard. 2006;20:41
137. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. Am J Health Syst Pharm. 2008;65(23):2276-84.
138. Boateng GO, Neilands TB, Frongillo EA, et al. Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. Frontiers in public health. 2018;6:149.
139. Newble D, Dawson B, Dauphinee D, et al. Guidelines for assessing clinical competence. Teaching and Learning in Medicine. 1994;6(3):213-20.
140. Norcini JJ, Blank LL, Arnold GK, Kimball HR. The mini-CEX (clinical evaluation exercise): a preliminary investigation. Annals of Internal Medicine. 1995;123(10):795-9.
141. Magnier KM, Dale VH, Pead MJ. Workplace-based assessment instruments in the health sciences. J Vet Med Educ. 2012;39(4):389-95.
142. Harden RM. Developments in outcome-based education. Med Teach. 2002;24(2):117-20.
143. Van Tartwijk J, Driessen EW. Portfolios for Assessment and Learning. Medical Teacher. 2009;31(9):790-801.
144. Segers M, Dochy F. New Assessment Forms in Problem-based Learning: The value-added of the students' perspective. Studies in Higher Education. 2001;26(3):327-43.

145. Baillie S, Warman S, Rhind S. A Guide to Assessment in Veterinary Medical Education: Royal College of Veterinary Surgeons Trust. Available from: http://www.bris.ac.uk/vetscience/media/docs/guide-to-assessment.pdf (2014, accessed 13 April 2022)
146. Harden RM, Stevenson M, Downie WW, et al. Assessment of clinical competence using objective structured examination. Br Med J. 1975;1(5955):447-51.
147. Smee S. Skill based assessment. BMJ (Clinical research ed). 2003;326(7391):703-6.
148. Pierre RB, Wierenga A, Barton M, et al. Student evaluation of an OSCE in paediatrics at the University of the West Indies, Jamaica. BMC Medical Education. 2004;4(1):22.
149. Solà-Pola M, Morin-Fraile V, Fabrellas-Padrés N, et al. The usefulness and acceptance of the OSCE in nursing schools. Nurse Education in Practice. 2020;43:102736.
150. Davis MH, Ponnamperuma GG, McAleer S, Dale VH. The Objective Structured Clinical Examination (OSCE) as a determinant of veterinary clinical skills. J Vet Med Educ. 2006;33(4):578-87.
151. Khan KZ, Gaunt K, Ramachandran S, et al. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: an historical and theoretical perspective. Med Teach. 2013;35(9):e1437-46.
152. Hecker K, Read EK, Vallevand A, et al. Assessment of first-year veterinary students' clinical skills using objective structured clinical examinations. J Vet Med Educ. 2010;37(4):395-402.
153. Read EK, Bell C, Rhind S, et al. The use of global rating scales for OSCEs in veterinary medicine. PloS one. 2015;10(3):e0121000.
154. Hodges BD. The Objective Structured Clinical Examination: three decades of development. J Vet Med Educ. 2006;33(4):571-7.
155. Boursicot K, Roberts T. How to set up an OSCE. The Clinical Teacher. 2005;2(1):16-20.
156. Bark H, Shahar R. The use of the Objective Structured Clinical Examination (OSCE) in small-animal internal medicine and surgery. J Vet Med Educ. 2006;33(4):588-92.
157. Bark H, Cohen R. Use of an objective, structured clinical examination as a component of the final-year examination in small animal internal medicine and surgery. J Am Vet Med Assoc. 2002;221(9):1262-5.
158. Bateman K, Menzies P, Sandals D, et al. Objective structured clinical examinations (OSCEs) as a summative evaluation tool in a ruminant health management rotation for final-year DVM students. J Vet Med Educ. 2008;35(3):382-8.

159. Annandale A, Fosgate GT, Bok H, et al. Ability of a bovine transrectal palpation objective structured clinical examination to predict veterinary students' pregnancy diagnosis accuracy. Veterinary Record. 2019;185(6):171-.
160. Annandale A, May CE, van der Leek ML, et al. Effect of a high-intensity one-week training programme and student-level variables on the bovine transrectal palpation and pregnancy diagnosis skills of final-year veterinary students. Vet Rec. 2020;187(11):e99.
161. Lumbis RH, Gregory SP, Baillie S. Evaluation of a Dental Model for Training Veterinary Students. J Vet Med Educ. 2012;39(2):128-35.
162. Eichel JC, Korb W, Schlenker A, et al. Evaluation of a training model to teach veterinary students a technique for injecting the jugular vein in horses. J Vet Med Educ. 2013;40(3):288-95.
163. Tan JY, Ma IWY, Hunt JA, et al. Video Recording in Veterinary Medicine OSCEs: Feasibility and Inter-rater Agreement between Live Performance Examiners and Video Recording Reviewing Examiners. J Vet Med Educ. 2021;48(4):485-91.
164. Artemiou E, Adams CL, Hecker KG, et al. Standardised clients as assessors in a veterinary communication OSCE: a reliability and validity study. Vet Rec. 2014;175(20):509.
165. Hecker KG, Adams CL, Coe JB. Assessment of first-year veterinary students' communication skills using an objective structured clinical examination: the importance of context. J Vet Med Educ. 2012;39(3):304-10.
166. Taylor A, Rymer J. The new MRCOG Objective Structured Clinical Examination--the examiners evaluation. J Obstet Gynaecol. 2001;21(2):103-6.
167. Norcini JJ. The Mini Clinical Evaluation Exercise (mini-CEX). The Clinical Teacher. 2005;2(1):25-30.
168. Faherty A, Counihan T, Kropmans T, et al. Inter-rater reliability in clinical assessments: do examiner pairings influence candidate ratings? BMC Medical Education. 2020;20(1):147.
169. Durning SJ, Cation LJ, Markert RJ, et al. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. Academic medicine : journal of the Association of American Medical Colleges. 2002;77(9):900-4.
170. Weller JM, Jolly B, Misur MP, et al. Mini-clinical evaluation exercise in anaesthesia training. British Journal of Anaesthesia. 2009;102(5):633-41.
171. Lörwald AC, Lahner FM, Nouns ZM, et al. The educational impact of Mini-Clinical Evaluation Exercise (Mini-CEX) and Direct Observation of Procedural Skills (DOPS)

and its association with implementation: A systematic review and meta-analysis. PloS one. 2018;13(6):e0198009.

172. Pelgrim EAM, Kramer AWM, Mokkink HGA, et al. In-training assessment using direct observation of single-patient encounters: a literature review. Advances in Health Sciences Education. 2011;16(1):131-42.
173. Norcini JJ, Blank LL, Duffy FD, et al. The mini-CEX: a method for assessing clinical skills. Ann Intern Med. 2003;138(6):476-81.
174. Alves de Lima A, Barrero C, Baratta S, et al. Validity, reliability, feasibility and satisfaction of the Mini-Clinical Evaluation Exercise (Mini-CEX) for cardiology residency training. Med Teach. 2007;29(8):785-90.
175. Wilkinson JR, Crossley JG, Wragg A, et al. Implementing workplace-based assessment across the medical specialties in the United Kingdom. Medical education. 2008;42(4):364-73.
176. Chang YC, Chen CK, Chen JC, et al. Implementation of the mini-clinical evaluation exercise in postgraduate Year 1 residency training in emergency medicine: Clinical experience at Chang Gung Memorial Hospital. Journal of Acute Medicine. 2013;3(3):110-5.
177. Cook DA, Beckman TJ, Mandrekar JN, et al. Internal structure of mini-CEX scores for internal medicine residents: factor analysis and generalizability. Advances in health sciences education : theory and practice. 2010;15(5):633-45.
178. Weijs CA, Coe JB, Hecker KG. Final-Year Students' and Clinical instructors' Experience of Workplace-Based Assessments Used in a Small-Animal Primary-Veterinary-Care Clinical Rotation. J Vet Med Educ. 2015;42(4):382-92.
179. L. Etheridge KB. Performance and workplace assessment. In: John A Dent RMH, Dan Hunt, editor. A Practical Guide for Medical Teachers. 5 ed. China: Elsevier; 2017. p. 267 - 73.
180. Prescott LE, Norcini JJ, McKinlay P, et al. Facing the challenges of competency-based assessment of postgraduate dental training: Longitudinal Evaluation of Performance (LEP). Medical education. 2002;36(1):92-7.
181. Cohen SN, Farrant PB, Taibjee SM. Assessing the assessments: U.K. dermatology trainees' views of the workplace assessment tools. Br J Dermatol. 2009;161(1):34-9.
182. Bell A, Hammond J. Directly Observed Procedural Skills – What do the Assessors Think? VetEd Symposium 2018 05-06 /07; Utrecht, The Netherlands.
183. Siau K, Crossley J, Dunckley P, et al. Direct observation of procedural skills (DOPS) assessment in diagnostic gastroscopy: nationwide evidence of validity and competency development during training. Surg Endosc. 2020;34(1):105-14.

184. Wood S, Baillie S, Warman S. Directly Observed Procedural Skill (DOPS) assessment of clinical examination skills in cattle: student perceptions. Cattle Practice. 2020;28:11-6.
185. Cobb KA, Brown G, Jaarsma DADC, et al. The educational impact of assessment: a comparison of DOPS and MCQs. Medical teacher. 2013;35(11):e1598-e607.
186. Laboratory Animal Science Association. Direct Observation of Practical/Procedural skills (DOPS) Available from: https://www.lasa.co.uk/dops/ (accessed 13 April 2022)
187. Laboratory Animal Science Association.Guiding Principles for Supervision and Assessment of Competence as required under EU and UK legislation. 2016.
188. Miller GE. The assessment of clinical skills/competence/performance. Academic medicine : journal of the Association of American Medical Colleges. 1990;65(9 Suppl):S63-7.
189. Hecker KG, Norris J, Coe JB. Workplace-based assessment in a primary-care setting. J Vet Med Educ. 2012;39(3):229-40.
190. Norcini J, McKinley DW. Concepts in assessment including standard setting. In: Dent JA, Harden RM, Hunt D (editors). A Practical Guide for Medical Teachers. 5 ed. China: Elsevier; 2017. p. 252- 9.
191. Livingston SA, Zieky MJ. Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests. Princeton: NJ: Educational Testing Service; 1982.
192. Cusimano MD. Standard setting in medical education. Academic medicine : journal of the Association of American Medical Colleges. 1996;71(10 Suppl):S112-20.
193. Kaufman DM, Mann KV, Muijtjens AM, et al. A comparison of standard-setting procedures for an OSCE in undergraduate medical education. Academic medicine : journal of the Association of American Medical Colleges. 2000;75(3):267-71.
194. Hejri SM, Jalili M. Standard setting in medical education: fundamental concepts and emerging challenges. Med J Islam Repub Iran. 2014;28:34.
195. Cizek GJ, Bunch MB. Standard setting: A guide to establishing and evaluating performance standards on tests. Thousand Oaks, CA: Sage Publications Ltd; 2007. xv, 352-xv, p.
196. Shepard L. Standard Setting Issues and Methods. Applied Psychological Measurement. 1980;4(4):447-67.
197. Medley DM. Teacher Competency Testing and the Teacher Educator. New Jersey: Ablex, Norwood; 1984. 51-94 p.

3. MATERIALS AND METHODS

MATERIAL AND METHODS

The papers presented in this thesis comprise different studies and this section presents an overview of the distinct material and methods used in the different studies.

In paper I (chapter 3), the acceptance of e-learning approaches by Laboratory Animal Science students was explored using a questionnaire administered to a total of 229 course participants. E-learning acceptance was assessed during 3 consecutive years (from 2012 to 2015) in a total of 15 Introductory courses (Directive Functions A and D *) and 6 Advanced courses (Directive Functions A, B and D*) .

In paper II (chapter 4) and paper III (chapter 5), the development and implementation, respectively, of a practical assessment for LAS context is presented. Paper II reports the development of a structured assessment rubrics to support the marking of participants, with the support of a core team constituted by professionals with expertise in teaching LAS and in biomedical teaching methods. In paper III (chapter 5), reports the implementation of *an* Objective Structured Laboratory Animal Science Exam (OSLASE) to assess researchers from different training courses. This study investigated particularly the exam reliability and trainees' OSLASE acceptability during approximately 3-year.

COURSES

The training courses in which the work presented in this thesis was situated, were held and organized at i3S-Institute for Research & Innovation in Health /IBMC- Institute of Molecular and Cell Biology, a biomedical research institution with a strong tradition of training, including a dedicated Advanced Training Unit. IBMC organizes training in laboratory animal science since 2005, with two course formats characterized as Introductory and Advanced. Introductory courses are directed at early stage researchers and were originally planned following the recommendations of former FELASA Category B (persons carrying out animal experiments) and later adapted to EU Directive 2010/63 Functions A and D. These courses are held 3 to 4 times per year and each course is spaced over time, with two to three months duration. Advanced courses target researchers who design and coordinate experiments with animals. Traditionally one course edition is organized per semester with an intensive program, 10 working days. In 2008 this course

(*) Article 23 of EU Directive 2010/63 established that staff working with animals shall be adequately educated according following functions:
(A) Carrying out procedures on animals
(B) Designing procedures and projects
(C) Taking care of animals
(D) Killing animals

was accredited by FELASA for category C (persons responsible for directing animal experiments), being this accreditation renewed in 2017 for the EU Directive 2010/63 Functions A, B and D, species-specific for Mouse, Rat, Fish and Zebrafish (FELASA course reference F020/08). Both training formats include a theoretical component, with classroom and online learning, practical sessions and evaluation. The Introductory course involves a period of tutor-led training, while the Advanced course also integrates group-assignment approaches. Paper I (Chapter 3) integrates a comparative table with description of the different programs of LAS training Courses in Materials and Methods section.

PARTICIPANTS

The participants enrolled in Laboratory Animal Science training between 2012 and 2021, in Introductory and in Advanced courses. When applicable, sample characterization is presented in each paper, separately.

METHODS

Paper I (chapter 3) | Students' acceptance of e-learning approaches in Laboratory Animal Science Training

Paper I explores the acceptance and perception of the use of e-learning in laboratory animal science training. For this study three questionnaires were specifically developed; (1) a Sociodemographic and Professional Questionnaire (SDPQ), (2) a Students' Expectations of E-learning in Laboratory Animal Science (SEELAS), and (3) a Questionnaire of E-learning Acceptance (QELA) as indicated in table 1.

General acceptance score was calculated using the sum of the five subscales items. The questionnaires described above are provided in Appendix I.

Paper II and III (chapter 4 and 5) | The methodology to develop a global rating scale for the assessment of researchers' competence and the implementation of OSLASE.

These two papers focused on the researchers' practical competence assessment. In the first study (paper II) the methodology to develop an assessment scale is presented with description of the different phases throughout the process until the final scale version. The final Global Rating Scales (GRSs) are provided in Appendix II.

The second study (III) addressed the implementation of the practical assessment - OSLASE. The exam was conducted in the context of introductory and advanced training during 3 consecutive years (2018 -2021). This study focused particularly on the analysis of interrater reliability and trainees OSLASE acceptability. Briefly, assessors' agreement was assessed through Weighted Kappa (Kw) and percentage of absolute agreement (% of agreement) after assessors conduct the exam, using the previously developed GRSs, to a minimum of 82 performances. This study analysed trainees experience using focus groups.

Table 1 | Questionnaire developed for Students' acceptance of e-learning approaches in Laboratory Animal Science Training (paper I)

	Characteristics	*Information collected*
SDPQ	5-item closed question	respondents' demographic information (gender, age, highest level of education completed, area of specialization, and current position/occupation.
SEELAS	7-item closed question	respondents' expectations of the use of e-learning at the start of the LAS course (e.g., difficulty, efficacy) and their previous experience with e-learning platforms.
QELA	32-item closed questions[3] likert-type scale comprises five subscales	Personal perception of the e-learning use Satisfaction with the e-learning contents Satisfaction with the e-learning organization Time management E-learning practical influence

DATA ANALYSIS

The statistical analyses performed for Paper I and Paper III were distinct, as detailed separately in each paper. The data analyses of students' acceptance (paper I) and inter-rater agreement (paper III) were conducted using IBM SPSS (25.0 version). Trainees' acceptability of the exam (paper III) was analysed using content analysis, with the software NVivo (QSR International 2022) version 12.

ETHICAL APPROVAL

Data collection with human participants (instrument application in the courses context and group interviews) were performed under approvals from the Joint Ethics Committee CHUP/ICBAS with the reference 2019/CE/P014(292/CETI/ICBAS) and 2021/CE/P013(P353/CETI/ICBAS).

[3] Regarding QELA questionnaire a 28-item version was administered to students in the introductory course, excluding four items related to the flipped classroom approach and thus not adjusted to that course format.

4. STUDENTS' ACCEPTANCE OF E-LEARNING APPROACHES IN LABORATORY ANIMAL SCIENCE TRAINING

Published in Laboratory Animals (2019), 54 (5), 487- 497

Original Article

Laboratory Animals Limited

Laboratory Animals
2020, Vol. 54(5) 487–497

DOI: 10.1177/0023677219879170
journals.sagepub.com/home/lan
SAGE

Students' acceptance of e-learning approaches in Laboratory Animal Science Training

Andreia Costa[1,2,3], Ana Costa[4] and I. Anna S. Olsson[1,2]

Abstract

Different online courses and training programs in Laboratory Animal Science (LAS) have emerged across Europe in recent years. E-learning appears to be a promising solution to achieve flexibility in training while meeting the quality criteria of demanding programs in short training periods. However, little is known about how students perceive e-learning in this context, and there is also a lack of specific and valid instruments to measure this perception. Within an exploratory study framework, the e-learning perception of 229 participants in 15 courses in Portugal using two different online training formats, flipped classroom and full online theoretical training, was assessed. For this purpose, the Questionnaire of E-learning Acceptance (QELA), a 32-item accordance Likert-type scale comprising five subscales was developed to explore the following: how participant perceive e-learning, satisfaction with organization and contents, perception of e-learning relevance for the time management, and its influence for practical training. In general, e-learning was well accepted and perceived to work well and be useful by the majority of courses participants, independently of the course level and e-learning format approach. These results indeed suggest that integration of e-learning is useful in LAS training. We also propose the QELA as a starting point for development and implementation of specific instruments to assess e-learning acceptance in LAS across a wider range of geographical and training contexts.

Keywords

teaching and training, e-learning, blended learning, flipped classroom, laboratory animal science training

Date received: 12 March 2019; accepted: 4 September 2019

Introduction

The traditional learning model based on classroom lectures has changed radically with the emergence of technology-supported learning approaches. Online education has expanded dramatically over the past decade,[1] and, as a consequence of the wide and fast development of internet access and multimedia technologies, e-learning is already considered a real alternative to face-to-face classes in many situations, redesigning through which channels, and at what time, knowledge is delivered.[2–4]

Conceptually, e-learning is the process of extending learning or delivering instructional materials by using telecommunication technology (such as internet, intranet/extranet, audio, video, satellite broadcast, interactive TV, and CD-ROM).[5] In the last 20 years, progress in the web and multimedia areas has fostered the growth of different teaching formats employing technology.[3,6] Within the wide concept of e-learning, more specific terms have emerged. Blended learning

[1]i3S, Instituto de Investigação e Inovação em Saúde, University of Porto, Portugal
[2]IBMC, Institute for Molecular and Cell Biology, University of Porto, Portugal
[3]ICBAS, Abel Salazar Institute of Biomedical Sciences, University of Porto, Portugal
[4]FPCEUP, Faculty of Psychology and Education Sciences of the University of Porto, Portugal

Corresponding author:
Andreia Costa, IBMC-Institute for Molecular and Cell Biology, Rua Alfredo Allen, 208, Porto 4200-135, Portugal.
Email: afcosta@ibmc.up.pt

refers to a specific format that combines elements from e-learning and traditional lectures, resulting in a hybrid/mixed learning system.[7] Flipped learning, or flipped classroom, refers to a format where students have the first contact with the didactic material outside the class, usually via reading or online presentations, and then use class time to do the harder work of knowledge assimilation, through problem solving, discussion, debates, etc.[8]

The term e-learning suggests a radically different type of learning. This is not the case: technology supported training relies on the same educational principles as more traditional formats.[9] E-learning is, however, particularly well suited to overcome some of the presently limiting issues of the traditional lecturing formats, such as time management difficulties originating in extensive and inflexible curricula where teachers and students face time constraints and schedule overlap. Unlike traditional lecture-centred models, where teachers and students need to be at the same place at the same time, e-learning makes it possible to train people anywhere and anytime, "on the go", while assuring that knowledge delivery is consistent.[4,10,11] The inclusion of an online component also promotes the active role of educators and learners, allowing teachers to be more than the content distributors, enhancing students' attitude, and promoting a dynamic and interactive learning.[12,13]

Against this background, the integration of e-learning seems to hold the potential to overcome many of the specific challenges of training in laboratory animal science (LAS). Specialized training in how to use animals in research is a legal obligation across Europe and in many countries worldwide.[14] In combination with increasing researcher mobility and decreasing availability of research support personnel, this presents challenges for both institutions and students. Research institutions have to make such training available frequently in order to enable new researchers to get started with their work.[15] Students on the other hand will have to make time for training in busy professional schedules. Indeed, e-learning is already considered the key to meet training requirements in a related situation, training of veterinary professionals, allowing users to manage their learning through the optimization of time availability and overcoming spatial constraints.[16,17] At our institution, we run two different training courses, each with its own particular instructional design challenges. The introductory course is for researchers and technicians about to start their work with animals. It is based on former (FELASA Category B) and present (Directive 2010/63/EU Functions A and D) recommendations for training for carrying out procedures on animals, and focuses on animal biology and experimental procedures. The advanced course has a wider coverage (based on former FELASA Category C recommendations and present Directive 2010/63/EU Functions A, B and D), targets more experienced researchers, and, in addition to biology and procedures, includes a large component on how to design and plan experiments with animals.[18–20]

Taking into account the training content and the educational level of students and teachers, LAS training seems particularly well suited for the variations on the flipped classroom format. Training is typically given in the context of second (MSc) and third (PhD) cycles of higher education, where students expect advanced-level training content and the possibility of interacting with teachers who are recognized experts in the respective discipline.[21] However, the nature of the course means that some rather basic content must also be learned.

Around 2010, we faced two distinct challenges for these two courses, both of which seemed possible to address by incorporating e-learning. For the introductory course, the challenge was to make the course available frequently enough that training could be made a requirement for access to the animal facility. For the advanced course, both students and teachers had difficulties making time for the 80 or more hours of training within the 2 weeks of the concentrated course. Of the many teaching methodologies using online approaches, the blended learning format seemed most appropriate. The combination of online teaching with classroom sessions, allows the delivery of more basic concepts before students have the opportunity to discuss and engage in face-to-face interactive activities to consolidate knowledge.[22–24] Successful results have been reported after e-learning implementation in veterinary education,[25,26] such as its use to complement ruminant endoscopy training,[22] farm animal clinical rotation,[23] or anatomical learning.[27] Moreover, research has highlighted the relationship between the flipped classroom format implementation and the use of online materials with better learning and performance.[28,29] This also includes practical skills, as demonstrated in a study where online learning resources were introduced to support the development of veterinary and animal science students' proficiency in safe and effective handling of livestock.[30]

An important aspect when evaluating teaching methods is to explore how they are perceived by students. A study investigating student perceptions of online lectures reported that the more satisfied the students are with online classes, the more the positive learning experience,[31] and a positive relation has also been found between the student satisfaction with e-learning, particularly the flipped learning format, and examination scores and performance.[32–34]

Globally, the present study investigated e-learning acceptance among participants in introductory and advanced training courses in LAS, using a crosssectional approach focusing on students after introducing the blended learning model. For this purpose, a questionnaire was developed on basis of existing instruments for e-learning evaluation.[2,35] In particular, the following different aspects of e-learning acceptance were studied: personal perception of the use; satisfaction with e-learning content, particularly with the online classes; satisfaction with e-learning organization and how the platform is organized concerning e-lectures, supplementary literature and other platform features; and the perception of time management and influence of the e-learning method on practical sessions.

Materials and methods

Type of courses

E-learning was implemented in both LAS training courses, as explained in detail in Table 1. For additional detail including distribution of content over e-learning versus classroom (see also Supplementary material Table S1).

E-learning platform

We implemented our learning environment based on the open source Moodle Learning Management System (LMS) version 2.6.11. Two separated e-learning course pages were developed at Moodle, one for the introductory and one for the advanced course. E-lessons were self-explanatory, including shorter videos, images, graphs and clinical cases and discussion/reflective slides (Figure 1). Formative and interactive quizzes were also included in order to facilitate the theory consolidation. The platform was further used as support throughout the courses for supplementary literature and students' forum.

The pedagogical approach integrating e-learning method was explained and clarified to the students before starting the course, by email and reinforced in the first lectures, and their e-learning participation was monitored over the training period. The students were free to work on e-learning when it suited them but had to comply with deadlines for completing it. In the introductory course category modules/e-lessons had to be completed by a certain deadline, at which time there was a final quiz. In the advanced course, deadlines were established to guarantee that students had completed the online content prior to the classroom session.

Instruments

Data collection comprised three questionnaires specifically developed for this study: a Sociodemographic and Professional Questionnaire (SDPQ), a Students' Expectations of E-learning in Laboratory Animal Science (SEELAS), and a Questionnaire of E-learning Acceptance (QELA):

(i) SDPQ is a 5-item closed question questionnaire that explores relevant data on the respondents' demographic information such as gender, age,

Table 1. Comparative description of the program of LAS training courses – Advanced and Introductory.

	Advanced Training	Introductory Training
Course Workload	Full-time use 10 full days	Part-time use 2 months
Organization	Theoretical component alternates with practical sessions	All theoretical component completed first, followed by practical sessions
E-learning theoretical approach	Flipped learning	All theoretical contents online
Theoretical program*	$\geq$40 h	$\geq$20 h
% of theoretical component delivered through e-learning*	30%	100%
Total course program	$\geq$80 h	$\geq$40 h
Course curricula | FELASA categories**	Category C – persons responsible for directing animal experiment	Category B – persons carrying out animal experiments

*Course duration are estimated values since online lessons have no strict time.
** The study covers a period before the functions based definition of training. Category C corresponds roughly to Functions A, B and D and Category B to Functions A and D identified in Directive 2010/63/EU.

Figure 1. Example of e-lessons slides, with illustrative images and explanatory videos.

highest level of education completed, area of specialization, and current position/occupation.

(ii) SEELAS is a 7-item closed question questionnaire that provides information about respondents' expectations of the use of e-learning at the start of the LAS course (e.g., difficulty, efficacy) and their previous experience with e-learning platforms.

(iii) QELA is a 32-item self-report measure that assesses students' acceptance of e-learning on the Moodle platform. QELA comprises five subscales: Personal perception of the e-learning use (PP; 10 items; *''The e-learning platform is user-friendly''; ''The e-learning platform works well''*), Satisfaction with the e-learning contents (SC; 7 items; *''Online quizzes helped me to learn effectively''; ''The e-lesson quizzes are easy to understand''*), Satisfaction with the e-learning organization (SO; 5 items; *''When I needed help to use the e-learning platform I knew where to find it''; ''The e-lessons are well structured and are easy to follow''*), Time management (TM; 5 items; *''The e-learning system enables me to invest more time in topics that interest me more''; ''The e-learning system enables me to control my learning progress, and invest more time in topics I find more difficult''*), E-learning practical influence (PI; 5 items; *''The videos shown in the e-learning helped me to prepare for practical classes of animal handling''; ''The videos shown in e-learning lectures helped me to deal with my possible animal fears''*), answered according to a 5-point Likert scale. General acceptance score was calculated using the items of the five subscales. In this study, this instrument showed positive correlations between the dimensions (between 0.28 and 0.49) and proved to have high internal consistency ($\alpha = 0.94$). The exploratory factorial analysis (EFA) extracted five factors consistent with the theoretical dimensions and explained 53.4% of variance. Globally, all the items presented good communalities and factor loading values.

A 28-item QELA version was administered to students in the introductory course, excluding four items related to the flipped classroom approach (one item of *Personal perception of the e-learning use*, and three of *Time management* dimension) and thus not applicable. Therefore, for any statistical analysis including this group of students only these 28 items were included.

Participants and procedure

The developed instruments were administrated to 229 LAS course participants in a total of 15 courses run between 2012 and 2015: 9 introductory (B) course editions, with a total of 90 participants mainly from our institution, and 6 advanced (C) course editions, with a total of 139 participants, with half of course participants originally from our institution and the other half from national and European research institutions.

Respondents were invited to participate in this study and completed the questionnaires in collective administrations on the last course day (after examination) in the presence of the researcher. Prior to each administration, the objectives of the study were explained to the participants, as well as the guarantee of anonymity and confidentiality, and that participation was voluntary. On average, it took 10 min to fill in the questionnaires.

Data analysis

To determine frequencies of e-learning users' satisfaction the global acceptance scale score and the subscales score were codified into three categories of satisfaction/usefulness (unsatisfactory, satisfactory, and very satisfactory).

The distribution of the total scale variable had a negative-skewness (left-skewed), with almost no respondents classifying e-learning acceptance as unsatisfactory/unhelpful; therefore, the percentile method to define groups was not applied. Considering the minimum and maximum possible total scores based in the 5 points answering items scale (1 = "Strongly disagree"; 2 = "disagree"; 3 = "neutral"; 4 = "agree"; and 5 = "Strongly agree"), the mean cutoff point dividing the negative and positive experience of users was established for the total scale and each dimension: Total acceptance scale (min:28; max:140; Mean:84), Personal perception (min:9; max:45; Mean:27), Satisfaction with contents (min:7; max:35; Mean:21), Satisfaction with organization (min:5; max.:25; Mean:15), Time management (min:5; max:25; Mean:15), and Practical influence (min:5; max:25; Mean:15). As the majority of the sample considered e-learning as satisfied/useful, the positive pole was also divided into satisfied/useful or very satisfied/very useful.

Inferential statistics were used to explore the overall acceptance of the e-learning use for different groups (course B and C). One-way ANOVA was the main statistical procedure to examine the effect of different sociodemographic, professional, and previous experience variables on acceptance issues: total acceptance score, personal perception, satisfaction with contents, satisfaction with e-learning, time management, and practical influence. Post hoc analyses with *Bonferroni* corrections were conducted to explore the differences between the subgroups. In all the statistical procedures, a p-level of 0.05 or below was considered to be significant. The data analyses were conducted using IBM SPSS (25.0 version).

Results

Sample characterization

Of the 229 respondents, 39.3% participated in the introductory (B) and 60.7% in the advanced (C) course. Respondent age ranged from 21 to 41 years old (mean (M) = 26.31; SD = 4.67) for the B course, and from 22 to 50 years old (M = 30.1; SD = 6.32) for the C course. In both courses, there was a female majority (B: 76.7%; C: 76.3%). Background training was primarily in biology (B: 50.0%; C:37.7%) and biochemistry (B:18.6%, C:23.2%). Highest level of education was, for the majority of B course participants, a master degree (58.4%) followed by a bachelor degree (24.7%), whereas among C course participants a master degree was most common (47.1%) followed by a PhD (37.0%). Professionally, the B course participants were PhD students (27.3%), undergraduate students (25.0%), or technicians (19.3%). C course participants were, as expected, at a different stage of career development and comprised PhD students (39.6%), post-doctoral fellows (34.3%) and principle investigators (PIs) (6.0%).

More than half of the total responders indicated no previous experience with e-learning resources (60.7%), with very similar distribution in the two training categories (B, 58.9%; C, 61.9%).

Concerning experience with laboratory animals, the majority of B participants indicated no experience (40.4%), while the majority of C participants indicated having more than 1 year of experience (47.8%). It was found that more respondents from the C course acknowledged a fear of laboratory animals (20.3%) than from the B course (11.1%) (Supplementary material Table S2).

Descriptive analysis

To assess students' e-learning global acceptance (QELA) and satisfaction with the different dimensions (personal perception of e-learning, satisfaction with contents, satisfaction with organization, time management, and practical influence), descriptive statistics were performed for all the participants and for each course group.

Globally, the general acceptance of the e-learning resources by LAS courses participants was very positive (62.1% classified as satisfactory and 30.4% as very satisfactory, with C course participants showing higher levels of satisfaction (94.2%) when compared with B participants (89.9%) (Figure 2).

Analyses of participants' personal perception of e-learning use showed high percentages of acceptance (C + B, 96.5%; C, 97.1%; B, 95.5%). Globally, and concerning satisfaction with the e-learning contents, course participants were satisfied (31,1%) and very satisfied (67.4%). Similar results were verified for e-learning organization, with courses participants considering it satisfactory (38.2%) and very satisfactory (60.1%).

The e-learning approach was considered useful (28.2%) and very useful (66.1%) concerning time control during the courses. Independently, B and C participants classified e-learning as useful (29.2% and 27.5%, respectively) for time management, while the majority of B and C participants considered it very useful (62.9% and 68.1%, respectively).

The results also showed that, globally, only a small percentage of both courses participants (4.4%) considered that e-learning was unhelpful and had no influence in practical sessions (B, 7.9% and C, 2.2%). The majority of participants (56.4%) classified the influence of the e-learning approach on practical sessions as helpful (B, 35.9% and C, 41.3%) or very helpful (B, 56.2% and C, 56.5%) (Figure 3).

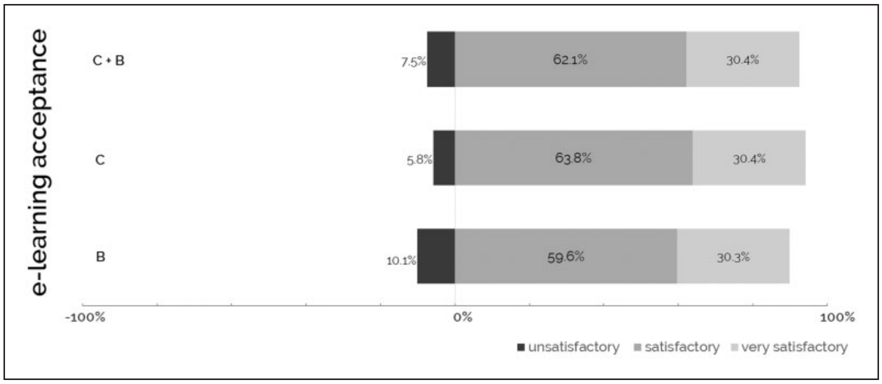

Figure 2. Acceptance of e-learning in LAS training courses by the participants of both courses (C + B). Participants of advanced (C) and introductory (B) courses categorized from very satisfactory to unsatisfactory and plotted as a diverging stacked bar chart.

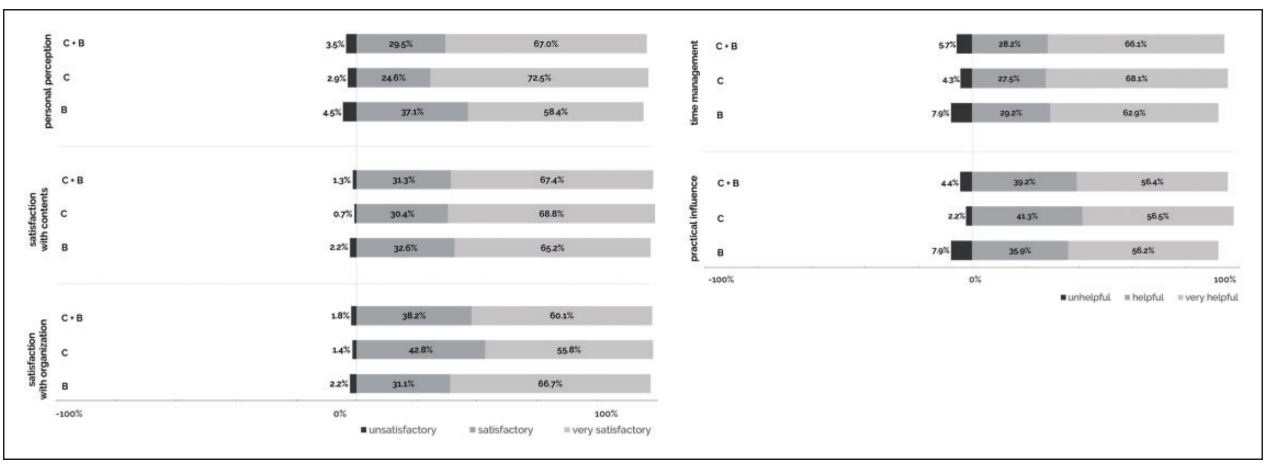

Figure 3. Analysis of participant acceptance of different e-learning features: personal perception of e-learning use, satisfaction with e-learning contents; satisfaction with organization; e-learning time management influence and e-learning practical training sessions influence. Bars represent answers from both course participants (C + B), advanced course participants (C), and introductory course participants (B), categorized from very satisfactory to unsatisfactory or very helpful to unhelpful, and plotted as a diverging stacked bar chart.

Differential analyses

General acceptance. Results showed that previous experience with laboratory animals affected general acceptance. Results indicated that advanced course participants with 1 year or more of previous experience with laboratory animals showed less acceptance of e-learning than students with no experience ($M_{none} = 129.05$, $SD = 11.84$; $M_{\geq 1year} = 118.43$, $SD = 18.03$; $F = 5.37$, $p < 0.01$). The same was verified for introductory course participants ($M_{none} = 108.50$, $SD = 12.07$; $M_{\geq 1year} = 95.75$, $SD = 28.69$; $F = 3.12$, $p = 0.05$). Gender, age, level of education, professional position, previous e-learning experience, and fear of laboratory animals did not influence the general acceptance of e-learning approach.

Personal perception of e-learning use. Previous experience with laboratory animals affected the C course participants' perception of e-learning use, with participants with no previous experience showing a more positive perception of e-learning when compared with participants with 1 year or more of experience ($M_{none} = 40.00$, $SD = 5.14$; $M_{\geq 1year} = 36.38$, $SD = 7.01$; $F = 3{,}97$, $p = 0.02$). Sociodemographic and academic variables did not affect participants' perception of e-learning use.

Satisfaction with e-learning content. For C course participants, satisfaction with the e-learning contents was affected by their position: PIs were more satisfied with content than technicians ($M_{PI} = 29.86$, $SD = 2.67$; $M_{Tech} = 24.00$, $SD = 4.57$; $F = 3.17$, $p = 0.02$). Also, specifically for C group students, previous experience with laboratory animals affected the satisfaction with e-learning content, revealing that participants with no experience were globally more satisfied than participants with 1 year or more of experience ($M_{none} = 28.79$, $SD = 2.75$; $M_{\geq 1year} = 26.18$, $SD = 3.97$; $F = 6.48$, $p < 0.01$).

Satisfaction with e-learning organization. Advanced course participants with more experience with laboratory animals were less satisfied with organization of e-learning. ($M_{none} = 19.92$, $SD = 2.85$; $M_{\geq 1year} = 18.02$ $SD = 3.25$; $F = 4.37, p = 0.01$).

Time management. None of the evaluated factors affected the different groups' satisfaction levels with time management.

E-learning practical classes' influence. Among C course participants, educational background affected how useful participants considered e-learning to be for practical training sessions. Participants coming from a biomedical sciences field consistently recognized a more positive influence of e-learning on practical classes than participants from biology ($M_{Biomed} = 20.81$, $SD = 2.77$; $M_{Biology} = 18.19$, $SD = 3.50$; $F = 4.48$, $p = 0.01$) and veterinary backgrounds ($M_{Biomed} = 20.81$, $SD = 2.77$; $M_{Vet} = 16.67$, $SD = 3.20$; $F = 4.48$, $p = 0.01$) (Supplementary material Table S3).

Discussion

The present study explored the general acceptance and satisfaction with the implementation of e-learning among participants in LAS training courses. In general, the results showed high levels of acceptance, both for the flipped classroom approach in the advanced training, and for the approach with all theoretical contents online that was used in the introductory training. Particularly, the results of the domains personal perception of e-learning use, satisfaction with content, and satisfaction with platform organization presented satisfaction percentages over 95.5% in the different courses. Our findings are consistent with several studies in other related fields reporting students' positive perception of e-learning resources.[24,25,36]

Time management is considered one of the most valuable aspects that e-leaning brings to educational context, allowing students to control their study time and content progression, and this was also recognized by the students in the present study, with 94.3% of the respondents classifying e-learning as useful for time management.

Another relevant aspect explored was the influence of e-learning in practical sessions, since e-lessons included didactic figures, schemes/graphics, videos of rodent behaviour, handling, and practical procedures. An overwhelming majority considered e-learning to be helpful/very helpful as a preparation for the practical classes. Indeed, the e-learning platform seems to have been perceived by students not only as a way to deliver theoretical content, but also as an instrument that influenced their practical learning positively. This fact may be related with the opportunity that online learning provides to learn and review the execution of practical procedures as many times as needed. Advanced course participants with a background in biomedical sciences had a more positive perception of e-learning as useful for practical classes as compared with participants with a veterinary or biology background, who, due to their previous academic training, are possibly more comfortable with the manipulation of living animals. Among the respondents, 32.6% had no previous experience and 16.7% indicated fear of working with laboratory animals; however, previous experience and fear did not affect how useful the e-learning was perceived to be for practical classes.

Indeed, acceptance of the e-learning approach seems to be highly influenced by whether students have previous experience with laboratory animals. Participants from the introductory (B) and the advanced (C) training showed the same consistent difference – participants with more experience (1 year or more) showed less acceptance of the e-learning approach than participants with no previous experience with laboratory animals. Similarly, and particularly in the C course students, previous experience affected personal perception of e-learning, and satisfaction with e-learning organization and e-learning content, with inexperienced students showing a more positive perception and higher levels of satisfaction. It is possible that this reflects an overall more positive attitude to laboratory animal science teaching among students who have less experience and hence may feel that they have more to gain from the course. Several factors may have contributed to this difference; factors such as motivation, age, and maturity have been reported to affect the learners' satisfactions with e-learning.[37,38] Additionally, it is possible that experienced participants also have less e-learning expectations, which can lead to low levels of satisfaction.[39,40]

Background also influenced the acceptance of the e-learning, with students with a biomedical sciences background showing greater e-learning acceptance than participants with a biology background. This is likely to be because students from biomedical areas are not entirely familiar with the contents addressed in the field of laboratory animal training, whereas participants from biology background are. The expectation and the novelty of the content may positively influence e-learning acceptance of this particular subgroup of participants, which have never studied similar subjects. This hypothesis is corroborated by the differences found in the satisfaction with e-learning content (higher in biomedical sciences than in students with biology background) and also the differences found in the positive e-learning influence on the practical components (higher in biomedical sciences than in students with a biology or veterinary background). The hypothesis that participants from a biomedical sciences background could be more experienced in e-learning was also explored but the results found did not corroborate this possibility.

Generally, our results show that acceptance of e-learning was not dependent of sociodemographic or academic variables such as gender, age, education, professional position, or previous e-learning experience.

This study explored the use of e-learning in two categories of training with different designs of e-learning use. Our results show that both e-learning formats, flipped and e-learning as the only vehicle for the theoretical component, are accepted in LAS courses. However, the results are limited, and more studies will be required to explore whether there is a preferable course design to include e-learning. Also, the fact that no group underwent classroom sessions only is a study limitation, because it limited the possible comparison between e-learning and classroom learning.

Measurement of psychometric properties such as perception and satisfaction requires appropriate instruments that are developed and adapted to the specific context of analysis and meet standard psychometric validity. Although exploratory, this study addresses the lack of measurement in the field and constitutes a precursory effort to develop a measure (QELA) to explore and assess student perceptions of the e-learning approach in a LAS context. The fact that the results distribution of the total acceptance variable presented a negative skewness, suggests that the instrument needs further adjustments in order to better reflect the less helpful or useful aspects of the e-learning approach, leading to the scale yielding overall very positive satisfaction outcomes. Further research should continue to develop, explore, and refine suitable measures to confirm the instruments' psychometric properties in different academic, professional, and cultural contexts in order to evaluate different e-learning strategies and their implementation in LAS training worldwide.

Conclusion

E-learning has been suggested as a useful resource to train people that carry out procedures or design experiments with laboratory animals. In this study we found e-learning to be well accepted by students in LAS courses at different training levels, who considered it enjoyable, useful, and effective in this specific context. Together with the finding that e-learning was accepted independently of sociodemographic, academic, and professional features, this suggests that incorporating e-learning into training programs is beneficial.

Acknowledgments

The authors would like to thank André Torres (IBMC / i3S) for the implementation and development of e-learning platform on open source Moodle Learning Management System and all course participants who collaborated in this study.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is a result of the project funded by Norte-01-0145-FEDER-000008 – Porto Neurosciences and Neurologic Disease Research Initiative at I3S, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (FEDER).

ORCID iDs

Andreia Costa https://orcid.org/0000-0001-8285-0065
I. Anna S. Olsson https://orcid.org/0000-0002-4369-8723

Supplemental Material

Supplemental material for this article is available online.

References

1. Kenzig MJ. Lost in Translation: Adapting a Face-to-Face Course Into an Online Learning Experience. *Health Promot Pract* 2015; 16: 625–628.
2. Wang YS, Wang HY and Shee DY. Measuring e-learning systems success in an organizational context: Scale development and validation. *Comput Hum Behav* 2007; 23: 1792–1808.
3. Zhang D, et al. Can e-learning replace classroom learning? *Commun ACM* 2004; 47: 75–79.
4. Ellaway R and Masters K. AMEE Guide 32: e-Learning in medical education Part 1: Learning, teaching and assessment. *Med Teach* 2008; 30(5): 455–473.
5. Holsapple CW and Lee-Post A. Defining, assessing, and promoting e-learning success: an information systems perspective. *Decision Sci J Innovative Educ* 2006; 4: 67–85.
6. Dalipi, F., et al., Rethinking the conventional learning paradigm towards MOOC based flipped classroom learning. 2017; DOI: 10.1109/ITHET.2017.8067791.

7. Graham CR. Blended learning systems: Definition, current trends, and future directions. In: Bonk CJ, and Graham CR (eds) The Handbook of Blended Learning: Global Perspectives, Local Designs. New York: Wiley, 2012.
8. Berrett D. How 'flipping' the classroom can improve the traditional lecture. *Chron Higher Educ* 2012; 31: A16–18.
9. Cook DA and McDonald FS. E-learning: is there anything special about the "E"? *Perspect Biol Med* 2008; 51: 5–21.
10. Govindasamy T. Successful implementation of e-Learning: Pedagogical considerations. *Internet Higher Educ* 2001; 4: 287–299.
11. Messaoudi T, et al. Evaluation of a new eLearning platform for distance teaching of microsurgery. *Chirurgie Main* 2015; 34: 109–112.
12. Ruiz JG, Mintzer MJ and Leipzig RM. The impact of E-learning in medical education. *Acad Med* 2006; 81: 207–212.
13. Lochner L, et al. Combining traditional anatomy lectures with e-learning activities: how do students perceive their learning experience? *Int J Med Educ* 2016; 7: 69–74.
14. European Commission, Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. Official Journal of the European Union: Brussels, 2010, p 33–79.
15. Crettaz von Roten F. Laboratory animal science course in Switzerland: participants' points of view and implications for organizers. *Lab Anim* 2018; 52: 69–78.
16. Alessandrini B, et al. Emergency management: e-learning as an immediate response to veterinary training needs. *Vet Ital* 2012; 48: 219–225.
17. Clausen PH, et al. Established and novel approaches for teaching and learning of veterinary parasitology in Berlin. *Vet Parasitol* 2018; 252: 58–61.
18. Wilson MS, Berge E, Maess J, et al., FELASA recommendations on the education and training of persons working with laboratory animals: categories A and C. Reports of the Federation of European Laboratory Animal Science Associations Working Group on Education accepted by the FELASA Board of Management. *Lab Anim* 1995; 29: 121–131.
19. Nevalainen T, et al., FELASA recommendations for the education and training of persons carrying out animal experiments (Category B). Report of the Federation of European Laboratory Animal Science Associations Working Group on Education of Persons Carrying out Animal Experiments (Category B) accepted by the FELASA Board of Management. *Lab Anim* 2000; 34: 229–235.
20. National Competent Authorities for the implementation of Directive 2010/63/EU on the protection of animals used for scientific purposes. A working document on the development of a common education and training framework to fulfil the requirements under the Directive. Brussels, 2019.
21. Gómez C. How e-learning is being used in veterinary training. *E-Learn Magazine: Competency-based Education* 2017; 15: 4–5.
22. Bernkopf M, Franz S and Baumgartner W. Experiences with a blended learning course for clinical veterinary education at the University of Veterinary Medicine Vienna, Austria. *Tieraerztliche Praxis Ausgabe Grosstiere Nutztiere* 2010; 38: 99–108.
23. Steele M, et al. Online tools for teaching evidence-based veterinary medicine. *J Vet Med Educ* 2013; 40: 272–277.
24. Borchers M, et al. Acceptance of case-based, interactive e-learning in veterinary medicine on the example of the CASUS system. *Tierarztl Prax Ausg K Kleintiere Heimtiere* 2010; 38: 379–388.
25. Gledhill L, et al. An international survey of veterinary students to assess their use of online learning resources. *J Vet Med Educ* 2017; 44: 692–703.
26. Xiberta P and Boada I. IVET, an interactive veterinary education tool. *J Anim Sci* 2019; 97: 932–944.
27. Adamczyk C, et al. Student learning preferences and the impact of a multimedia learning tool in the dissection course at the University of Munich. *Ann Anat* 2009; 191: 339–348.
28. Khan S, et al. Video projected practical examination as an introduction to formative assessment tool for undergraduate examination in pathology. *Indian J Pathol Microbiol* 2019; 62: 79–83.
29. O'Flaherty J and Phillips C. The use of flipped classrooms in higher education: a scoping review. *Internet Higher Educ* 2015; 25: 85–95.
30. Klupiec C, et al. Development and evaluation of online video teaching resources to enhance student knowledge of livestock handling. *Aust Vet J* 2014; 92: 235–239.
31. Scagnoli NI, Choo J and Tian J. Students' insights on the use of video lectures in online classes. *Br J Educ Technol* 2019; 50: 399–414.
32. A. Petrarca, C., et al. Evaluation of eLearning for the teaching of undergraduate ophthalmology at medical school: a randomised controlled crossover study. *Eye* 2018; 32: 1498–1503.
33. Dooley LM, et al. Implementing the flipped classroom in a veterinary pre-clinical science course: student engagement, performance, and satisfaction. *J Vet Med Educ* 2018; 45: 195–203.
34. García-Iglesias MJ, et al. Mixed-method tutoring support improves learning outcomes of veterinary students in basic subjects. *BMC Vet Res* 2018; 14: 35.
35. Ozkan S and Koseler R. Multi-dimensional students' evaluation of e-learning systems in the higher education context: an empirical investigation. *Comput Educ* 2009; 53: 1285–1296.
36. Bartova E and Halova D. E-learning for veterinary students. In: Houska M, Krejci I and Flegl M (eds) Efficiency and responsibility in education. Prague 6: Czech University Life Sciences Prague, 2014, pp. 27–31.
37. Piccoli G, Ahmad R and Ives B. Web-based virtual learning environments: a research framework and a preliminary assessment of effectiveness in basic IT skills training. *MIS Quarterly* 2001; 25: 401–426.
38. Hong K-S. Relationships between student' and instructional variables with satisfaction and learning from a Web-based course. *Internet Higher Educ* 2002; 5: 267–281.

39. Paechter M, Maier B and Macher D. Students' expectations of, and experiences in e-learning: Their relation to learning achievements and course satisfaction. *Comput Educ* 2010; 54(1): 222–229.

40. Francis Umbit A. The effects of expectations and satisfaction towards e-learning among students. *J Modern Educ Rev* 2016; 6: 603–611.

Résumé

Différents cours et programmes de formation en ligne ont émergé partout en Europe ces dernières années en ce qui concerne les sciences des animaux de laboratoire (LAS). La formation en ligne semble constituer une solution prometteuse pour développer la flexibilité des programmes de formation tout en répondant aux critères de qualité exigeants de programmes de formation courts. On ne sait toutefois que peu de choses sur la façon dont les étudiants perçoivent l'apprentissage en ligne dans ce contexte, et il existe peu d'instruments valables et spécifiques pour mesurer cette perception.

Dans un cadre d'étude exploratoire, la perception de l'apprentissage en ligne de 229 participants à 15 cours se déroulant au Portugal en utilisant deux différents formats de formation en ligne, à savoir en classe inversée et en formation théorique en ligne complète, a été évaluée. À cette fin, le questionnaire d'acceptation de l'apprentissage en ligne (QELA), qui comprend 32 éléments de type Likert et 5 sous-échelles, a été développé afin d'explorer la perception de la formation en ligne par les participants, la satisfaction vis-à-vis de l'organisation et du contenu, la perception de la pertinence de la formation en ligne en ce qui concerne la gestion du temps et son influence sur la formation pratique.

En général, l'apprentissage en ligne était bien accepté et perçu comme fonctionnant bien, et comme étant utile par la majorité des participants aux cours, indépendamment du niveau du cours et de l'approche adoptée quant à la formation en ligne. Ces résultats suggèrent en effet que l'intégration de l'apprentissage en ligne s'avère utile en ce qui concerne la formation aux LAS. Nous proposons également le QELA comme point de départ pour le développement et la mise en œuvre d'instruments spécifiques permettant d'évaluer l'acceptation de l'apprentissage en ligne en ce qui concerne les LAS à travers un large éventail de régions géographiques et de contextes de formation.

Abstract

In den vergangenen Jahren sind überall in Europa Online-Kurse und Ausbildungsprogramme für Versuchstierkunde (LAS) erarbeitet worden. E-Learning scheint eine vielversprechende Lösung zu sein, um Flexibilität in den Trainingsprogrammen zu erreichen und gleichzeitig die Qualitätskriterien anspruchsvoller Programme in kurzen Ausbildungszeiten zu erfüllen. Es ist jedoch wenig darüber bekannt, wie die Studierenden E-Learning in diesem Zusammenhang wahrnehmen, und es fehlt zudem an konkreten und validen Instrumenten zur Messung dieser Wahrnehmung.

Im Rahmen einer explorativen Studie wurde die E-Learning-Wahrnehmung von 229 Teilnehmern in 15 Kursen in Portugal unter Einsatz von zwei verschiedenen Online-Trainingsformaten, dem Umgedrehten Unterricht und dem kompletten theoretischen Online-Training, bewertet. Zu diesem Zweck wurde der Fragebogen zur Akzeptanz von E-Learning (QELA), eine 32 Punkte umfassende Likert-artige Skala mit 5 Subskalen, entwickelt, um Folgendes zu untersuchen: wie die Teilnehmer E-Learning wahrnehmen, die Zufriedenheit mit Durchführung und Inhalt, die Wahrnehmung der E-Learning-Relevanz für das Zeitmanagement und ihr Einfluss auf die praktische Ausbildung.

Grundsätzlich wurde E-Learning von der Mehrheit der Kursteilnehmer gut angenommen und als gut funktionierend und nützlich empfunden, unabhängig von der Kursstufe und dem Konzept des E-Learning-Formats. Diese Ergebnisse legen durchaus nahe, dass die Integration von E-Learning in die LAS-Schulung sinnvoll ist. Wir empfehlen den QELA auch als Grundlage für Entwicklung und Umsetzung konkreter Instrumente zur Bewertung der Akzeptanz von E-Learning in LAS in einem breiteren Spektrum von geografischen und Ausbildungskontexten.

Resumen

En los últimos años, han surgido distintos cursos en línea y programas de formación sobre ciencias de animales de laboratorio (LAS) en Europa. El aprendizaje en línea parece ser una solución prometedora para conseguir flexibilidad en los programas de formación cumpliendo a su vez los criterios de calidad de

distintos programas exigentes en periodos de formación breves. Sin embargo, no hay mucha información sobre cómo los estudiantes perciben la formación en línea en este contexto y también existe una falta de instrumentos válidos y específicos para evaluar esta percepción.

Dentro de un marco de estudio exploratorio, se evaluó la percepción de la formación en línea de 229 participantes en 15 cursos en Portugal que utilizaron dos formatos de formación en línea diferentes, aula invertida y formación teórica en línea total. Para este fin, el Cuestionario de Aceptación de Formación en Línea (QELA), una escala de tipo Likert de 32 elementos de conformidad que comprendía 5 subescalas, fue creado para explorar: cómo los participantes perciben el aprendizaje en línea, la satisfacción con la organización y contenido, la percepción de la relevancia del aprendizaje en línea para la gestión de tiempo y su influencia para la formación práctica.

En general, el aprendizaje en línea fue bien aceptado y se consideraba satisfactorio y útil por la mayoría de los participantes de cursos, independientemente del nivel del curso y del método del formato del aprendizaje en línea. Estos resultados, en efecto, sugieren que la integración de la formación en línea es de gran utilidad para la formación LAS. También proponemos el QELA como punto de partida para desarrollar e implementar instrumentos específicos para evaluar la aceptación del aprendizaje en línea en LAS en una amplia gama de contextos geográficos y de formación.

5. THE ASSESSMENT OF RESEARCHERS’ COMPETENCE IN EXPERIMENTAL PROCEDURES WITH LABORATORY ANIMALS: A THREE-STEP METHODOLOGY TO DEVELOP A GLOBAL RATING SCALE

Published in Laboratory Animals (2021), 55 (5), 463-471

Original Article

Laboratory Animals Limited

Laboratory Animals
0(0) 1–8

DOI: 10.1177/00236772211017767
journals.sagepub.com/home/lan
SAGE

The assessment of researchers' competence in experimental procedures with laboratory animals: A three-step methodology to develop a global rating scale

Andreia Costa[1,2,3], **Sofia Lamas**[1,2], **Manuel J Costa**[4] and **I Anna S Olsson**[1,2]

Abstract

To conduct animal experiments, researchers must be competent to handle and perform interventions on living animals in compliance with regulations. Laboratory animal science training programmes and licensing bodies therefore need to be able to reliably ensure and certify the professional competence of researchers and technicians. This requires access to assessment strategies which can verify knowledge as well as capturing performative and behavioural dimensions of assessment. In this paper, we describe the process of developing different global rating scales measuring candidates' competence in a performative assessment. We set out the following sequence, with three crucial phases, in the process of scale development: (a) Item Development, (b) Scale Development and (c) Piloting of the Scale. We note each phase's different sub-steps. Despite the emergent need to ensure the competence of researchers using animals in scientific procedures, to our best knowledge there are very few species and procedure/skill specific assessment tools for this purpose, and the assessment methodology literature in the field is very limited. This paper provides guidance for those who need to develop and assess proficiency in laboratory animal procedures by setting out a method that can be used to create the required tools and illustrating how competence assessment strategies can be implemented.

Keywords

Assessment of competence, professional competence, global rating scales and checklists, laboratory animal science proficiency, laboratory animal training supervision

Date received: 6 November 2020; accepted: 26 April 2021

Introduction

Researchers and technicians conducting animal experiments must be professionally competent to handle and perform interventions on living animals in compliance with regulations and the 3Rs principle.[1] Laboratory animal science (LAS) training programmes, licensing bodies and research institutions are therefore expected to reliably ensure and certify the professional competence of researchers for such purposes.[1,2] Building upon the definition provided by Epstein and Hundert (2002) for medical doctors,[3] we define professional competence as the habitual use of knowledge, technical

[1]i3S – Instituto de Investigação e Inovação em Saúde, University of Porto, Portugal
[2]IBMC – Institute for Molecular and Cell Biology, University of Porto, Portugal
[3]ICBAS – Abel Salazar Institute of Biomedical Sciences, University of Porto, Portugal
[4]ICVS – Life and Health Sciences Research Institute, School of Medicine, University of Minho, Portugal

Corresponding author:
Andreia Costa, University of Porto Institute for Molecular and Cell Biology, Rua Alfredo Allen, 208, Porto, Porto 4200-135, Portugal.
Email: afcosta@ibmc.up.pt

skills, emotions, values and reflection in daily practice, by researchers handling animals for the sake of animal welfare and scientific validity. Any demonstration of professional competence will involve a complex assessment with cognitive, psychomotor and emotional dimensions. This calls for the development of fit-for-purpose assessment tools capable of differentiating competent animal scientists from their less well prepared, insufficiently competent fellow professionals. In this endeavour, it is helpful to examine research that has solved contextually identical challenges in other fields, in particular human medicine, in which there is a longer tradition of developing assessment tools for interventions involving living beings.

The level of credibility of an assessment instrument at the performative and behavioural levels will show how accurately it captures the various dimensions of competence of those being assessed. The term *validity* is used to refer to this accuracy of measurement. The development of new instruments, with the ultimate aim of achieving reliable and valid evaluations, involves a series of steps. The demonstration of an instrument's validity requires evidence-based argument that the instrument captures what it was intended to.[4,5]

Professional competence to perform experiments with animals is a multidimensional construct. Three dimensions can be separated: (1) theoretical and technical knowledge (e.g. of maximum volumes of solution per administration, the selection of an appropriate route and/or site); (2) technical ability (e.g. in animal restraint or needle/syringe manipulation); and (3) attitudinal ability (including motivation, professionalism, values, attitude and empathy when handling an animal). Systematic assessment of professional competence therefore needs to go beyond tests in the knowledge dimension. It must use tools to assess skills and behaviours, whether in controlled or in real, professional settings. In designing these complex assessments it is useful to consider a simple framework developed for medical education. 'Miller's pyramid'[6] classifies competence at four levels: first, there is (a) knows and (b) knows how (both cognition). Then there is (c) shows how and (d) does (performative and behavioural). At cognition levels (a) and (b) competence is usually assessed through essays or written tests designed with close-ended or selected response questions (see Khan and Aljarallah[7] and Hawkins and Swanson[8]). At higher levels (c) and (d), behaviour is assessed through observation of performance in professional tasks and simulated or authentic settings. Performance is generally marked on checklists or global rating scales.[9,10] Simulation-based assessments in controlled settings are easier to implement than workplace-based assessments and also less influenced by assessor bias.[11,12] An example of a widely established simulation method in health sciences and veterinary education is the Objective Structured Clinical Examination (OSCE).[13,14] This versatile,[15] valid, reliable and feasible[16] approach to the assessment of clinical skills at the third level of Miller's pyramid has been widely adopted across distinct educational areas in many countries.[14] OSCEs involve a circuit of short steps, or stations, at each of which the candidate is required to perform a specific task, generally under examiner supervision (so-called manned stations). Candidates rotate through a specific circuit, completing all the stations.[17,18]

The creation of an assessment in LAS inspired in the OSCE model – for example, an Objective Structured Laboratory Animal Science Exam (OSLASE) – judging the actual technical performance of trainees would make an important contribution to the evaluation of professional competence in animal science. A representative OSLASE will ideally include a variety of stations so that it covers the full range of learning outcomes for the training. If we are to develop a complete OSCE we will need to determine which stations to include. It will also be necessary to develop scales to capture candidate performance at each station. Although there should be specific scales for each station, all of the scales can be designed and evaluated following the same methodology.

In this paper, we set out a method for developing global rating scales. We illustrate the process of instrument development, focusing on a tool for the assessment of applicants' proficiency in the execution of key procedures with laboratory rodents. Whereas we present this approach for an OSLASE, the methodology we present is not limited to this specific context. It is relevant for the development of a variety of instruments for assessing professional competence in handling and performing procedures with live animals (or even in dummies, when restrictions exist) – whether the assessment is part of an examination after a period of training or workplace-based and designed for new researchers arriving at animal facilities with previous experience and training who are required to demonstrate their competence.

Compared with related fields such as medicine and veterinary medicine, the literature on assessment methodology for LAS is limited. The importance of assessment strategies to verify competence has been highlighted in several papers and reports,[2,19–22] but to our knowledge the only publicly available assessment tool is the set of competence assessment templates published by The Laboratory Animal Science Association.[21] The Education and Training Platform for Laboratory Animal Science have been working through their working groups aiming to develop Directly Observed Practical Skills for objective competence assessment.[22] The work presented in this paper, the first in the field, outlines a methodology integrating the initial and basic steps needed to generate a global rating scale for the assessment of professional competence in frequently used procedures with live animals. Our work is guided by recommendations made by Boateng and colleagues for the systematic and sequential process of developing a scale through Item Development, Scale Development and Piloting of the Scale.[23]

The development of a scale to assess competence in handling and procedures involving living animals

A. Item Development

The Item Development phase (A in Figure 1) includes procedure selection and content validity analysis. In other words, it involves selection of the task(s) for the candidates to perform and of the aspects of performance to be assessed. Procedures were selected meeting the following criteria: (a) an essential procedure related to work with living rodents which authorized researchers must perform without supervision; (b) to some extent applicable across the most common laboratory rodents, implying the same basic principles for both rats and mice (e.g. subcutaneous administration is performed in a similar manner); and (c) pervasive in training and certification programmes across countries. These criteria can be adjusted according to the trainer's and institution's needs, and their establishment prior to item development is relevant for instrument validity.

Items were generated using a deductive method, starting with the revision of reference materials[23,24] and breaking the procedure down into key, independent sub-steps. This technique was repeated for each procedure separately, resulting in one scale for each of the four selected procedures.

The focus of the next step was on content validity,[25] the ability of the instrument to measure the competence of researchers in handling and performing procedures in rodents. Work here was led by a core team containing professionals with expertise in LAS (teaching as well as the performance of procedures) and experts with know-how in biomedical teaching methods. The team comprised the four co-authors of this paper, all of whom participated in, or oversaw, all steps of the method's development. The team worked together to define descriptor words, to ensure specificity, and to select appropriate vocabulary to represent a specific movement. As a component of content validity, we also considered face validity (i.e. the degree to which assessors judge that the scale items are appropriate to the targeted assessment objectives[24]) by discussing the scales with two independent experts with experience in performing and teaching the selected procedures. The output of this phase was a first version of the scale consisting of a list of aspects of performance to be assessed and descriptors of performance at different levels of competence.

B. Scale Development

The second phase, Scale Development (B in Figure 1), consisted in scale pre-testing and administration, and item reduction. During pre-testing, scale descriptors were refined by the team to ensure that they were meaningful to the assessors (our target population). This is crucial to guarantee that the descriptors underpinning the object of judgement – the sub-step of the skill, such as 'depth of puncture/angle' during the intraperitoneal administration – make sense for the assessors that are going apply the instrument. Misinterpretations and errors were corrected before the scales were administered with the desired effect. During the administration phase, the scales were tested under both simulated (i.e. with volunteers performing on mannequins) and authentic conditions (i.e. competent staff executing the procedures at the facility). Pretesting and administration were carried out in different phases by two of the panel members, AC and SL (the assessors, both LAS tutors), and with a panel of professionals, all with expertise in the procedures under assessment with the global rating scale (GRS).

The focus at this stage was on establishing consensus on the approval criteria and the standard for the procedure (i.e. a description of what performing the procedure correctly involves).

Setting the expected performances for each rating category of the procedure helped, at this stage, to clarify which items were essential (sometimes referred to as 'killing items' because they were items in which poor performance would always lead to the failing of a candidate). This reduced the scale so that it included only critical items for performance assessment (item reduction). Item descriptors were also condensed where it was suitable to do so, and open-ended items were completely removed from the instrument, since they turned out to be too time-consuming to fit into the assessment of a short procedure. Team discussion during standard setting was also important in establishing the score categories for performance rating. One of the main refinements introduced in this phase was the reduction of the number of rating categories used to score the instrument items. We concluded that three categories were adequate given the procedure's duration (short procedures < 90 seconds), so that this reduction could be made without compromising the

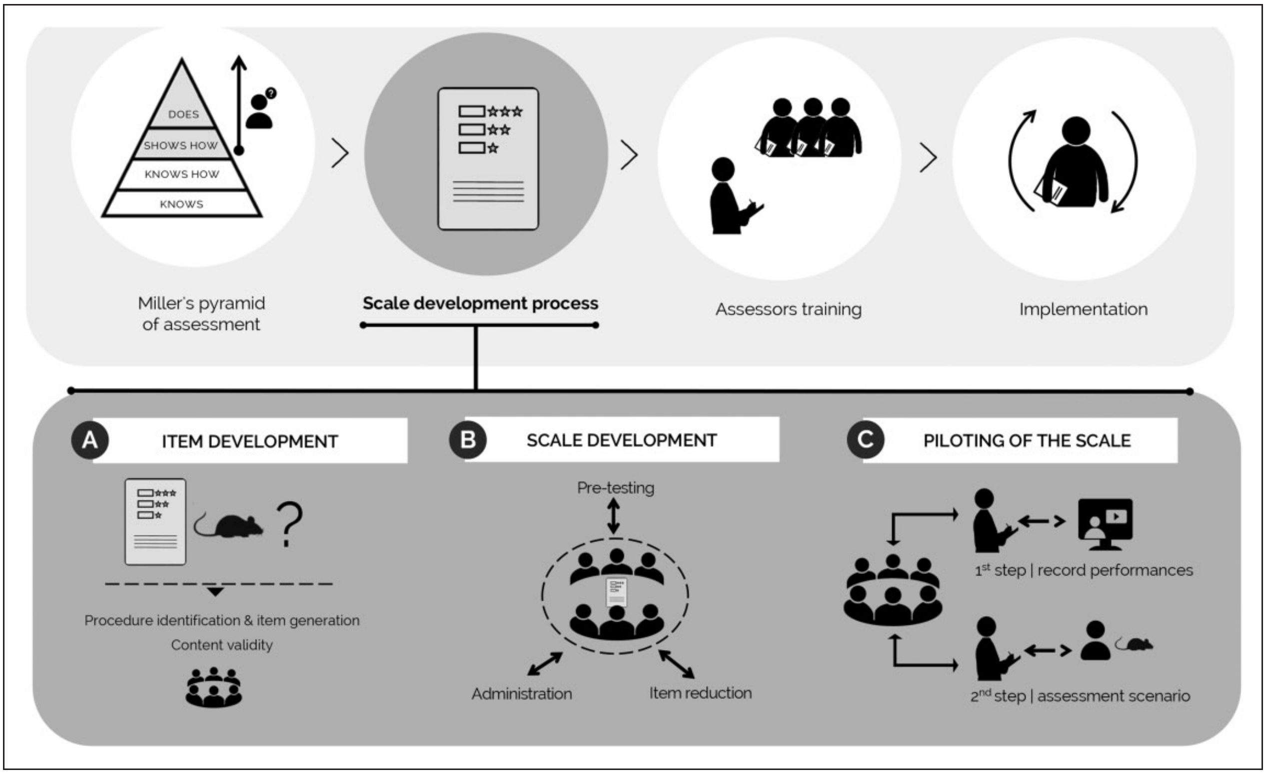

Figure 1. Scheme illustrating the operationalization of Miller's pyramid of assessment to evaluate professional competence in laboratory rodent experimentation at the performative and behavioural levels. The figure highlights the most representative stages of scale development: A. Item Development, B. Scale Development and C. Piloting of the Scale.

assessment process. We recommend that the time candidates are allowed to attempt the procedure is limited to 90 seconds in order not to compromise animal welfare. A trainee who is proficient to work without supervision needs around 30 seconds to execute the task.

This stage involved a highly iterative process, with sequential feedback from the assessors to the other team members (after each administration) and vice-versa (restoring the main adjustments), and with rotation between the three stage-steps whenever necessary, until the stage had been completed. This resulted in a complete version of the scale, which, as soon as it was working acceptably well, was applied individually by two assessors who had participated in the development (AC and SL). After establishing that the assessments were concordant and gave similar GRS outputs when applied to the same demonstrations, we moved to the third and final step in the development of the scale.

C. Piloting of the Scale

Piloting of the Scale (C in Figure 1) was the last stage and served to ensure the scale's applicability with a wider range of assessors. Before introducing the scale into assessments involving living animals, the first pilot trial used video recordings of comparable procedures on mannequins, performed at distinct levels of proficiency. Videos of excellent, satisfactory and poor performances were captured at our institute animal facilities, or obtained from free online Internet resources, to mimic different types of possible candidate performance.

When the scale had been trialled, and after assessors had been appropriately trained using video recordings in performance rating, the scale was finally introduced in a real evaluation scenario with course participants. All participants involved gave prior consent to the inclusion of their rating performances in this study. The scale was administered several times, in different contexts, with feedback from assessors to the team and subsequent amendments of the scale where necessary. For details of these scale adjustments, see Figure 2.

Discussion and conclusion

EU directive 2010/63/EU and its corresponding implementation guidance documents[2] require the competency of members of the animal users community to be assessed. This study explores the methodology behind the development of a GRS to assess competence, as

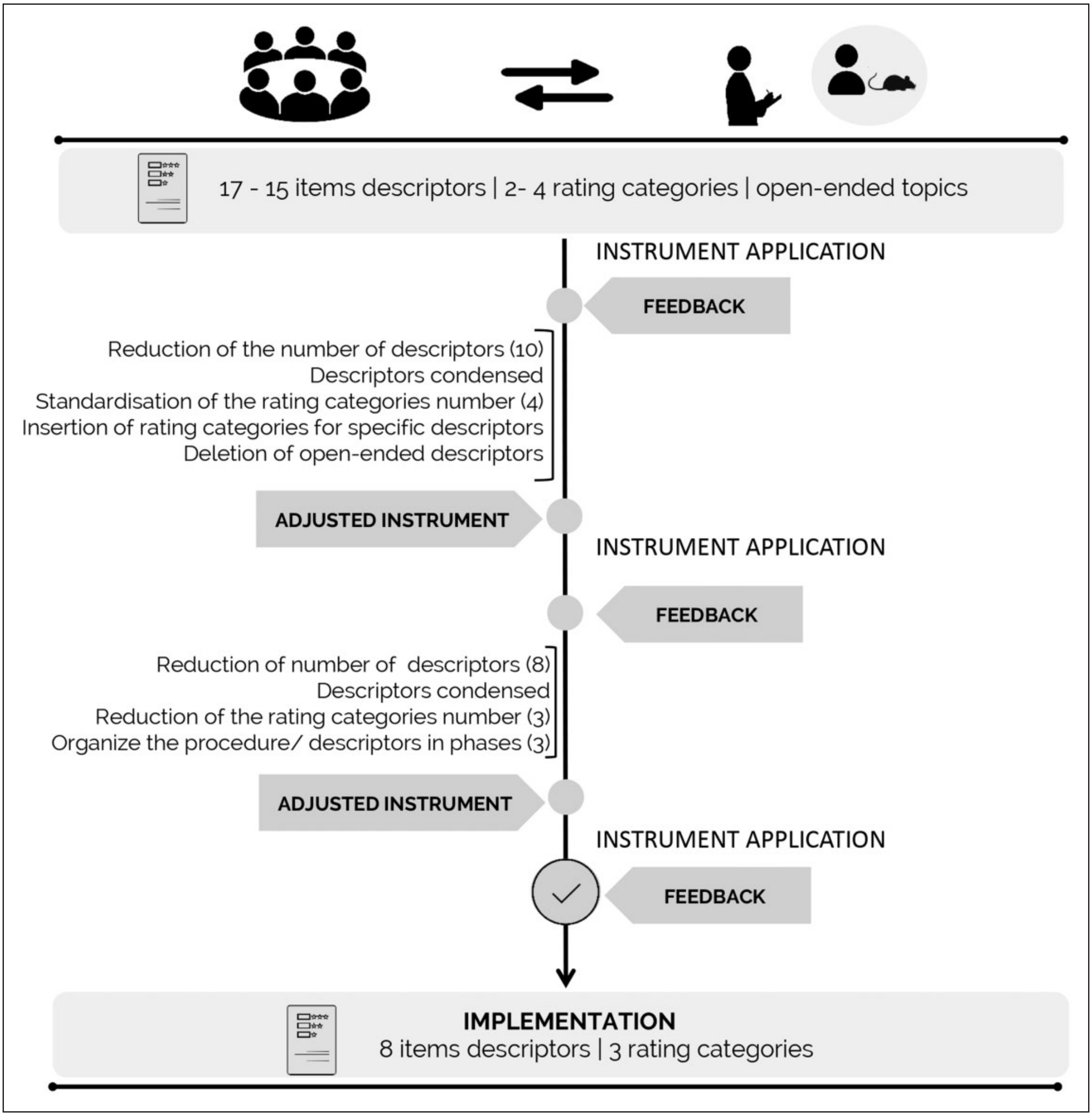

Figure 2. Steps in the adjustment of the instrument mainly resulting from phase C. Piloting of the Scale. The definitions of educational terms are provided by the National Council on Measurement in Education: https://www.ncme.org/resources/glossary.

current literature provides minimal guidance on how these instruments should be developed. Although there is currently very little research into education methods, competency assessment and proficiency verification in the LAS field, much can be learned from the field of medical education. In this paper, we have described the development of a rating instrument for assessing competence in conducting procedures with live animals (or in dummies). Our aim was to illustrate a general method of scale development that can be used to respond to the challenges presented by the EU directive. The method was adopted from established procedures for the development of OSCE stations employed in medical education,[26] drawing also on Boateng and colleagues' systematic process for instrument development,[23] and is applicable to several contexts where assessment of professional competence is required.

In the present work, we used simulations for scale development and assessor training. In our approach, in its final implementation the scale will be used in the real context of procedures on living animals. This is in line with the guidance on competency assessment given under Directive 2010/63/EU, which states that 'The assessor should observe and evaluate the trainee performing the procedures to assess practical competence'.[2] The final GRSs applying the methodology described in this paper are available for review and use.[27]

The methodology we have reported ensures validity in the development process. Validity, that is, how accurately a measurement or result corresponds to what it is intended to measure,[7] is key in the development of assessment instruments. Downing has highlighted that 'validity is the sine qua non of assessment, as without evidence of validity, assessments in medical education have little or no intrinsic meaning'.[8] The overall validity of an exam for assessment of competence, whether an OSCE or other type of practical assessment, is directly influenced by the validity and reliability of the tools and process of scoring examinees. The scoring tools can be checklists or GRSs. Checklists are easy to apply and show higher inter-rater reliability, while GRSs are capable of differentiated levels of proficiency and expertise.[30–32] We opted for the GRS in our approach, as this best fitted our aim to develop an instrument that will be applicable in training scenarios, where capturing increasing levels of proficiency while limiting the number of performance assessments with live animals is important. GRSs were fit for purpose assessment tools for summative – 'assessment of learning' – and for formative – 'assessment for learning' – purposes. We have primarily used the GRSs in exams to infer whether candidates had developed competence to a level suitable for being 'signed off' to work unsupervised. The instruments are equally useful to provide objective and directed feedback and orientate the trainee process of skills acquisition. For example, the assessment records of individual candidates identify the specific aspects that candidates need to develop further. The process described in this paper is applicable in the development of a checklist as well as a GRS.

Professional competence to perform experiments with animals is a multidimensional construct, encompassing theoretical and technical knowledge, technical ability and attitudinal competence. Our focus was on the dimensions of technical and attitudinal ability. However, the systematic verification of the scales' content validity also indicated that the instruments we developed were suitable for the procedure under analysis and adequate to measure the candidate's proficiency in those dimensions – technical and attitudinal – through evaluation of the proposed task.

Taking all scale developmental adjustments into account, we found that reducing the number of items and condensing the descriptors to bring the amount of text down to a minimum were absolutely key to guaranteeing that the scale is feasible, as assessors struggled to read lengthier definitions while supervising and scoring simultaneously. In practice, the procedure described here took place over a period of several months. However, the timeframe needed will be highly dependent of the resources available at the institution, particularly the availability of persons with expertise (on the targeted competencies/procedure or skill to be assessed) and the opportunities for testing the scales in assessment situations. Based on this experience, in the present paper we address how this process can be simplified by following the systematic and standard approach we propose here. Particular attention should be paid to the choice of items and descriptors. The number of items should be limited (< 10 per scale), preference should be given to 'killing items' and redundant items avoided. Descriptors should be short and objective, and the items and descriptors listed chronologically according to the procedure. Although time will be required to prepare the assessors for the scale administration, by following this approach the scale development will be less time consuming.

The present paper focuses on the scale development process. After finalizing and validating the scale, and before implementing it as an assessment tool in an institutional setting, the assessors must be trained. As in every observational examination, the reliability of assessors will condition the reliability of summative assessments of experimental procedures with laboratory animals with the GRS. Indeed, under the same conditions for the observed competence in examinations, all assessors should interpret the evidence collected with the GRSs and reach the same decision on the competence of a candidate. Therefore, the assessors need training to ensure reliability and minimize errors likely caused by dissimilar scale use.[33,34] Training should ensure that assessors reliably use the instruments to identify correct and incorrect performances. Assessors need clear instructions on how to apply the scale and they need to be aware of pass/fail criteria. In the present context, where living animals are used, it is also crucial for assessors to know how, and when, to interrupt demonstrations to avoid unnecessary animal distress. The training could consist of iteration cycles of performance ratings of video recordings of performance at several levels of competency, followed by debriefing, until the necessary level of accuracy is attained. There is a lack of guidance and research on methodology to develop assessment tools for use in the LAS context with established validity, despite the emergent need to ensure the competence of researchers who use animals in scientific procedures. In the present paper, we have outlined, and illustrated, a method that we hope will provide guidance for those involved in the development of further assessment tools. Our aim has been to support the specific scientific community of LAS researchers in their efforts to implement valid assessment strategies centring on the verification of professional competence during initial training stages or later.

Acknowledgements

The authors would like to thank the researchers and technicians who agreed to participate in the expert panel, the final training candidates who collaborated in this study and the i3S communication unit for its support with the illustrations for this paper.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: this work is a result of the project funded by Norte01-0145-FEDER-000008 – Porto Neurosciences and Neurologic Disease Research Initiative at i3S, supported by Norte Portugal Regional Operational Programme (NORTE2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (FEDER).

ORCID iDs

Andreia Costa https://orcid.org/0000-0001-8285-0065
Sofia Lamas https://orcid.org/0000-0002-1665-1989
I Anna S Olsson https://orcid.org/0000-0002-4369-8723

References

1. European Parliament and Council of the European Commission Union. *Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes*. Strasbourg, France: Council of Europe, 2010.
2. European Commission guidance document. National Competent Authorities for the implementation of Directive 2010/63/EU on the protection of animals used for scientific purposes. A working document on the development of a common education and training framework to fulfil the requirements under the Directive. Brussels, http://ec.europa.eu/environment/chemicals/lab_animals/pdf/guidance/education_training/en.pdf (2014, accessed 13 April 2020).
3. Epstein RM and Hundert EM. Defining and assessing professional competence. *JAMA* 2002; 287: 226–235.
4. Sullivan GM. A primer on the validity of assessment instruments. *J Grad Med Educ* 2011; 3: 119–120.
5. Kimberlin CL and Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm* 2008; 65: 2276–2284.
6. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990; 65(9 Suppl.): S63–S67.
7. Khan MU and Aljarallah BM. Evaluation of Modified Essay Questions (MEQ) and Multiple Choice Questions (MCQ) as a tool for Assessing the Cognitive Skills of Undergraduate Medical Students. *Int J Health Sci* 2011; 5(1): 39–43.
8. Hawkins RE and Swanson DB. Using written examinations to assess medical knowledge and its application. In: Holmboe ES and Hawkins RE (eds). *Practical guide to the evaluation of clinical competence*. Philadelphia, PA: Mosby, Elsevier, 2008, pp. 42–59.
9. Norcini J, Anderson B, Bollela V, et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 *Conference*. *Med Teach* 2011; 33: 206–214.
10. Norcini J, Anderson MB, Bollela V, et al. Consensus framework for good assessment. *Med Teach* 2018; 40: 1102–1109.
11. Brydges R, Hatala R, Zendejas B, et al. Linking simulation-based educational assessments and patient-related outcomes: A systematic review and meta-analysis. *Acad Med* 2015; 90: 246–256.
12. Weersink K, Hall AK, Rich J, et al. Simulation versus real-world performance: A direct comparison of emergency medicine resident resuscitation entrustment scoring. *Adv Simul (Lond)* 2019; 4: 9.
13. Davis MH, Ponnamperuma GG, McAleer S, et al. The Objective Structured Clinical Examination (OSCE) as a determinant of veterinary clinical skills. *J Vet Med Educ* 2006; 33: 578–587.
14. Khan KZ, Ramachandran S, Gaunt K, et al. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: An historical and theoretical perspective. *Med Teach* 2013; 35: e1437–e1446.
15. Harden RM, Stevenson M, Downie WW, et al. Assessment of clinical competence using objective structured examination. *Br Med J* 1975; 1: 447–451.
16. Patricio MF, Juliao M, Fareleira F, et al. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? *Med Teach* 2013; 35: 503–514.
17. Boursicot K and Roberts T. How to set up an OSCE. *Clin Teach* 2005; 2: 16–20.
18. Agarwal A, Batra B, Sood A, et al. Objective structured clinical examination in radiology. *Indian J Radiol Imaging* 2010; 20: 83–88.
19. Clifford P, Melfi N, Bogdanske J, et al. Assessment of proficiency and competency in laboratory animal biomethodologies. *J Am Assoc Lab Anim Sci* 2013; 52: 711–716.
20. Hansen AK and Sørensen DB. Strategies for the assessment of competence in laboratory animal science courses. *Lab Anim (NY)* 2014; 43: 359–363.
21. Jenning M and Berdoy M. Guiding principles for supervision and assessment of competence as required under

EU and UK legislation 2nd ed. Report, LASA Education – Training and Ethics Section, July 2016.
22. Education and Training Platform for Laboratory Animal Science. DOPS and SOP, https://etplas.eu/education/dops-and-sop/ (accessed 1 April 2021).
23. Boateng GO, Neilands TB, Frongillo EA, et al. Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Front Public Health* 2018; 6:149.
24. Zamanzadeh V, Ghahramanian A, Rassouli M, et al. Design and implementation content validity study: Development of an instrument for measuring patient-centered communication. *J Caring Sci* 2015; 4: 165–178.
25. Haynes SN, Richard DCS and Kubany ES. Content validity in psychological assessment: A functional approach to concepts and methods. *Psychol Assess* 1995; 7: 238–247.
26. Zayyan M. Objective structured clinical examination: The assessment of choice. *Oman Med J* 2011; 26: 219–222.
27. Costa A, Lamas S, Costa MJ et al. Global rating scales to assess the competence of researchers in handling and performing substance administrations in laboratory rodents. Repositório Aberto da Universidade do Porto, https://hdl.handle.net/10216/133419 (2021, accessed 20 April 2021).
28. Downing SM and Haladyna TM. Validity threats: Overcoming interference with proposed interpretations of assessment data. *Med Educ* 2004; 38: 327–333.
29. Downing SM. Validity: On meaningful interpretation of assessment data. *Med Educ* 2003; 37: 830–837.
30. Regehr G, MacRae H, Reznick RK, et al. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med* 1998; 73: 993–997.
31. Ringsted C, Ostergaard D, Ravn L, et al. A feasibility study comparing checklists and global rating forms to assess resident performance in clinical skills. *Med Teach* 2003; 25: 654–658.
32. Ilgen JS, Ma IW, Hatala R, et al. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Med Educ* 2015; 49: 161–173.
33. Yule S, Rowley D, Flin R, et al. Experience matters: Comparing novice and expert ratings of non-technical skills using the NOTSS system. *ANZ J Surg* 2009; 79: 154–160.
34. Feldman M, Lazzara EH, Vanderbilt AA, et al. Rater training to support high-stakes simulation-based assessments. *J Contin Educ Health Prof* 2012; 32: 279–286.

L'évaluation de la compétence des chercheurs en procédures expérimentales avec des animaux de laboratoire: méthodologie en trois étapes pour développer une échelle de notation globale.

Résumé

Afin de mener des expériences sur des animaux, les chercheurs doivent être compétents pour manipuler et effectuer des interventions sur des animaux vivants, conformément à la réglementation. Les programmes de formation en sciences animales de laboratoire et les organismes de délivrance de licences doivent donc être en mesure d'assurer et de certifier de manière fiable la compétence professionnelle des chercheurs et des techniciens. Cela nécessite l'accès à des stratégies d'évaluation à même de vérifier les connaissances et de saisir les dimensions performatives et comportementales de l'évaluation. Dans cet article, nous décrivons le processus d'élaboration de différentes échelles de notation mondiales mesurant la compétence des candidats dans une évaluation performative. Nous avons défini la séquence suivante, avec trois phases cruciales, dans le processus de développement de l'échelle: (A) Développement de point, (B) Développement de l'échelle et (C) Utilisation de l'échelle. Nous notons les différentes sous-étapes de chaque phase. Malgré le besoin émergent d'assurer la compétence des chercheurs utilisant les animaux dans les procédures scientifiques, à notre connaissance il existe très peu d'outils d'évaluation spécifiques aux espèces et aux procédures/compétences à cette fin, et la littérature sur la méthodologie d'évaluation dans ce domaine est très limitée. Cet article fournit des conseils à ceux qui ont besoin de développer et d'évaluer la compétence dans les procédures relatives aux animaux de laboratoire en définissant une méthode qui peut être utilisée pour créer les outils requis et en illustrant comment les stratégies d'évaluation des compétences peuvent être mises en œuvre.

Die Bewertung der Kompetenz von Forschern bei experimentellen Verfahren mit Labortieren – dreistufige Methodik zur Entwicklung einer umfassenden Bewertungsskala

Abstract

Forscher müssen über entsprechende Kompetenz verfügen, wenn sie Tierversuche durchführen, das heißt, sie müssen wissen, wie mit lebenden Tieren umzugehen ist und wie Eingriffe unter Einhaltung der Vorschriften erfolgen müssen. Ausbildungsprogramme für Versuchstierkunde und Zulassungsstellen müssen daher in der Lage sein, die fachliche Kompetenz von Forschern und Technikern zuverlässig zu gewährleisten und zu zertifizieren. Dies erfordert den Zugang zu Beurteilungs-Strategien, die sowohl Wissen verifizieren als auch leistungs- und verhaltensbezogene Dimensionen der Beurteilung erfassen können. Im vorliegenden Beitrag beschreiben wir den Prozess der Entwicklung verschiedener umfassender Bewertungsskalen, die die Kompetenz der Kandidaten in einer leistungsbezogenen Beurteilung messen. Wir stellen den folgenden Ablauf mit drei entscheidenden Phasen im Prozess der Skalenentwicklung dar: (A) Itementwicklung, (B) Skalenentwicklung und (C) Pilotierung der Skala. Wir vermerken die einzelnen Teilschritte jeder Phase. Trotz der sich abzeichnenden Notwendigkeit, die Kompetenz von Forschern sicherzustellen, die Tiere in wissenschaftlichen Verfahren einsetzen, gibt es unseres Wissens nur sehr wenige arten- und verfahrens- bzw. fähigkeitsspezifische Beurteilungsinstrumente für diesen Zweck, und die Literatur zur Beurteilungsmethodik auf diesem Gebiet ist sehr beschränkt. Dieser Artikel bietet eine Anleitung für jene, die Kompetenz in Verfahren mit Labortieren entwickeln und beurteilen müssen, indem er eine Methode aufzeigt, die zur Erstellung der erforderlichen Instrumente verwendet werden kann, und veranschaulicht, wie Kompetenzbeurteilungsstrategien implementiert werden können.

Evaluación de la competencia de investigadores de procesos experimentales con animales de laboratorio: Un modelo de tres pasos para crear una escala de calificación global

Resumen

Para realizar experimentos con animales, los investigadores deben ser lo suficientemente competentes para manipular y llevar a cabo intervenciones con animales vivos en cumplimiento de las normativas vigentes. Por tanto los programas de formación científica con animales de laboratorio así como los organismos de licencias deben ser capaces de garantizar y certificar razonablemente la competencia profesional de los investigadores y técnicos. Esto requiere acceso a estrategias de evaluación para poder verificar el conocimiento del que estos disponen así como registrar las dimensiones de comportamiento y ejecución de las evaluaciones. En este estudio, describimos el proceso de creación de distintas escalas de calificación global para medir la competencia de los candidatos mediante una evaluación de carácter performativo. Establecemos la siguiente secuencia, con tres fases cruciales, en el proceso de creación de la escala:(A) Desarrollo del elemento, (B) Creación de la escala y (C) Pilotaje de la Escala. Registramos los distintos subpasos de cada fase. A pesar de la necesidad emergente de garantizar la competencia de los investigadores que utilizan animales para procesos científicos, según nuestro leal saber y entender, existen pocas herramientas de evaluación concretas de procesos/pericia y de pocos tipos para este fin, y el número de documentos sobre metodologías de evaluación en este sector es muy limitado. Este estudio ofrece unas directrices para aquellos que necesitan desarrollar y evaluar la eficacia de los procesos con animales de laboratorio estableciendo un método que puede utilizarse para crear las herramientas adecuadas y para ilustrar cómo pueden implementarse estrategias de evaluación de la competencia.

6. AN OBJECTIVE STRUCTURED LABORATORY ANIMAL SCIENCE EXAMINATION (OSLASE) TO ENSURE RESEARCHER'S PROFESSIONAL COMPETENCE IN LAS

Submitted to Special Edition of Laboratory Animals Journal on Education

Published Manuscript: Appendix VI

AN OBJECTIVE STRUCTURED LABORATORY ANIMAL SCIENCE EXAMINATION (OSLASE) TO ENSURE RESEARCHER'S PROFESSIONAL COMPETENCE IN LAS

Andreia Costa [1,2,3*], Sofia Lamas [1,2], Maria Rui Correia [1,2], Maria Salomé Gomes [1,2,3], Manuel João Costa [4], I. Anna S. Olsson [1,2]

[1] i3S, Instituto de Investigação e Inovação em Saúde, University of Porto, Portugal
[2] IBMC, Institute for Molecular and Cell Biology, University of Porto, Portugal
[3] ICBAS, Abel Salazar Institute of Biomedical Sciences, University of Porto, Portugal
[4] ICVS, Life and Health Sciences Research Institute, School of Medicine, University of Minho, Portugal

Manuscript submitted

Abstract

Evaluating the competence of personnel working with laboratory animals is currently a challenge. Directive 2010/63/EU establishes that staff must have demonstrated competence before they perform unsupervised work with living animals. Nevertheless, there is a lack of research into education and training in Laboratory Animal Science (LAS) and the establishment of assessment strategies to confirm researchers' competence remains largely unexplored.

In this study we analyze the implementation of a practical assessment strategy, using the Objective Structured Laboratory Animal Science Exam (OSLASE) developed by us in previous work to assess professional competence. The interrater reliability (IRR) was determined based on the assessors rating of candidates' performance at different OSLASE stations using Weighed Kappa (Kw) and percentage of agreement. Focus group interviews were also conducted to access trainees' acceptability regarding the OSLASE.

There was a moderate to good Kw for the majority of the scales' items ($0.79 \pm 0.20 \leq Kw \geq 0.45 \pm 0.13$). Percentages of agreement also revealed acceptable values (≥ 75%) for all scales' items, excluding one. Trainees reported the OSLASE to have a positive impact on engagement during practical training, and that it clarified the standards established for the performance and the skills a trainee would need to improve. These preliminary results illustrate how assessment strategies, such as OSLASE, can be implemented in a way that is useful for both assessors and trainees.

Keywords

competence and proficiency; performance assessment; Objective Structured Laboratory Animal Science Exam (OSLASE)*; LAS* Assessment impacts; Education and Training in Laboratory Animal Science

Introduction

Driven by an increased awareness of the importance of ensuring researchers' professional competence in animal experimentation to safeguard animal welfare, the interest in the design and implementation of innovative assessment methods is increasing notoriously. This appears not only as an imposition of the revised European legislation (1, 2), but also as a way to guarantee the 3Rs application in practice (3).

The Directive 2010/63/EU establishes that staff working with living animals must be "adequately educated and trained" and "supervised in the performance of their tasks until they have demonstrated the requisite competence" (2). The related working document for a common education and training framework further clarifies that assessment of practical competence should rely on observations of trainees performing the procedures (1). Such regulations call out for more attention of education and training supervisors to competence and assessment constructs in the Laboratory Animal Science (LAS) context (4). Ensuring professional competence in LAS through observational assessments is critical at different stages of learning. For example, as a final examination at the conclusion of LAS training before the beginning of autonomous work, or as a formative work-integrated assessment as part of Continuing Professional Development (CPD).

The field of laboratory animal science can draw on the extensive knowledge on the assessment of professional competence developed in medical and health professions education. In these fields, driven by specific professional regulations and the agenda of patient safety, concepts as competence, performance and assessment methods validity and reliability have been largely explored (5-7). Assessments are crucial elements in the educational process and have a role to play in every phase of professional development (8). Assessment plays a key role in the process of learning, having a positive impact and influencing learners in how they plan and carry out their studies (6-9).

Regardless of when the practical assessment takes place, choosing an adequate format is a key element of ensuring that an assessment is capable of measuring what is intended. In LAS, there is a need for the observational measurement of professional competence, by capturing dimensions of performance other than cognitive knowledge, which is suitably assessed by traditional written tests. In this sense, the format and the design of assessments must be planned to be able to capture psychomotor (technical ability) and if possible emotional/attitudinal aspects (such as professionalism or empathy with animals (10). Although LAS training programs are increasingly focusing on professional competence in practice, assessments focused on abilities in this domain are still lacking. As part of an ongoing project *at* our institution aimed at developing an Objective Structured Laboratory Animal Science Exam (OSLASE) of researchers' competence, we have previously developed instruments to capture the performance of researchers in handling and performing substance administrations in laboratory rodents (10, 11). In this study we continue the development process evaluating the Interrater Reliability of assessors and the acceptability of the OSLASE for those who have taken the examination. These are two important elements to better understand the relevance and the assessment impact. Interrater reliability is critical for the reliability of any assessment. As the OSLASE is quite different from existing examinations, understanding how participants experience the examination should reveal the impact on their motivation and engagement, exam preparation and development of competence.

Materials and methods

Study design and setting

This study included participants who completed Advanced (Directive Functions A, B and D) and Introductory (Directive Functions A and D) LAS training for rodent species. The Objective Structured Laboratory Animal Science Exam (OSLASE) practical assessment is an examination designed to test LAS skills performance and competence inspired in the widely used model of the OSCE (Objective Structured Clinical Examination) in health professions education (10, 12). The exam consisted of four manned stations in which participants were observed performing essential and routine procedures with laboratory rodents: mouse restraint (MR), rat restraint (RR), mouse intraperitoneal administration (MIA) and rat subcutaneous administrations (RSA). All stations were timed and consisted on handling and restraint of a live rodent (MR; RR) and execution of procedure with mannequins (MIA; RSA). In both contexts (Introductory and Advanced courses), performances were assessed by trained examiners supported by a global rating scale (GRS) specifically developed for these procedures (11).

In our institution, achieving passing mark in the OSLASE (practical assessment) is necessary to successfully conclude the LAS course and those who fail need to retake the practical training. In the introductory course, trainees take the OSLASE and a written assessment at the end of training. These are completed with a period of tutor-supervised training at the animal facility. A choice of dates for the OSLASE is offered between two options spaced approximately by one-month. In the advanced course, OSLASE was integrated as a performance evaluation in the rodent handling practical module, which has two sessions and the trainee gets to choose in which session he/she wants to do the assessment. In this case, OSLASE consists in a demonstration of mouse and rat restraint. In both training levels, trainees are assessed individually in a standardized scenario by a trained assessor. Table 1 summarizes the main aspects of each of the training courses.

Table 1 | Training courses assessment aspects.

	INTRODUCTORY course	ADVANCED course
ASSESSMENT TIMING	Final [At the end of practical training]	During training process [Between practical course modules, end of handling module]
OSLASE STATIONS	mouse restraint rat restraint mouse intraperitoneal administration rat subcutaneous administration	mouse restraint rat restraint
DATA COLLECTED	interrater agreement trainees exam acceptability	trainees exam acceptability

OSLASE set up

In each station, assessors used a specific global rating composed of 7 to 9 items with descriptors, listed sequentially. The final item for each scale is the "Global Impression" item. All scales' items present 3 score categories for performance rating: fail, pass, clear pass.

In introductory LAS courses, candidates were assessed individually, by the first and second author of this study, at the end of the training period. The two assessors assessed each candidate (N=92) independently, at the same time, by direct observation of performances. Not all candidates included in the sample performed the complete exam, including the 4 stations. Some candidates fail at the first item of the scales which leads to failing the performance (ex: fail to pick animal from the cage). In these circumstances, the exam is ended. Also, a few candidates only received training for one of the species and therefore performed only stations related with that species (e. g. when training for work with mouse, performed MR and IPM).

Evaluation of Interrater agreement

The evaluation of the OSLASE for the purpose of this paper included assessing the agreement between the ratings of the two assessors - the interrater reliability (IRR) - in two ways: the percentage of absolute agreement (% of agreement) and weighted kappa (Kw) statistic. These are recommended methods to determine assessors agreement, due to the recognition of the importance of decisions in healthcare and clinical research supported by interrater reliability results (13). Percent of absolute agreement between the two assessors was the number of concordant scores divided by the total number of scores. A range from 75% to 90% of absolute agreement is considered to be at an acceptable level (13, 14). The weighted kappa (Kw), with linear weighting, was calculated for all scale items of the four global rating scales and for the global impression item (the last item of each scale). Kw allows to weight disagreements differently, which is especially useful in studies applying scales with ordered rating categories, so that it confers greater emphasis to large differences between ratings than to small differences (15-17). Kappa values representing the agreement between raters were interpreted according to the following: <0.20 poor ; 0.21− 0.40 fair; 0.41 − 0.60 moderate; 0.61 − 0.80 good and 0.81 − 1.00 very good (18) . The analyses were performed using the statistical software package SPSS 21, SPSS Inc. Chicago, USA.

Evaluation of acceptability

To capture a variety of perceptions in both course levels concerning the acceptability of OSLASE, 13 laboratory animal science course trainees (7 participants from the advanced course (2 different editions) and the remaining from the introductory (3 different courses)) participated in focus group interviews. Eligible trainees had successfully completed the course in either level (Introductory and Advanced) in the preceding 24 months and had received the certificate. Participants were invited by email. All the respondents which accepted the invitation were integrated in the interviews. They were previously informed about the study aim and conditions and gave their consent to participate. The choice to conduct focus groups had the vision of delving more deeply into students' views. The interview guide was developed by the study authors and adapted to the different courses (see table 1). Because of Covid-19 restrictions, interviews were carried out online (Zoom). There were a total of 4 focus group interviews conducted during a 7 months period. Trainees were allocated to focus groups according to course type (two focus groups per type) and availability. The interviews lasted approximately 90 minutes, with an initial introduction followed by a moderator (last author) facilitated discussion, whereas an assistant moderator (third author) took notes during interviews. Neither the moderator nor the assistant moderator were related with the OSLASE administration. Participants were first provided an introduction about the purpose of the study, and then script questions followed. Participants freely discussed their own experiences, in combination with prompts when necessary. With the participants' permission, all sessions were audio and video recorded and the content was transcribed verbatim. Participants identification was pseudonymized. The verbatim transcripts were coded using the software NVivo (QSR International 2022) version 12 for qualitative analysis. The first and the last author decided on a first set of codes and coded each transcription separately, before meeting to confirm the coding and revise the citations. At the end, 5 categories were identified (Choice of timing to perform the exam; Effect on trainee practical preparation before perform OSLASE, Experience and perception, OSLASE impact on trainee learning process and OSLASE relevance and importance).

Results

Determining agreement between raters

The mouse restraint scale items related to “tail grasp attempts”, “tail hold” and “suspension duration” showed high percentages of agreement between raters (≥ 85%) but weighted kappa (Kw) revealed poor agreement, lower than 0.40. The remaining items of this scale showed moderate to good values of Kw (0.45 ±0.13 ≤ Kw ≥ 0.61±0.08). The lowest percentage of agreement for all the items of this scale was displayed for the “global impression” item (71%) although it presented a moderate value of Kw (0.51 ±0.07) (Table 2 – A).

Regarding the rat restraint scale, the Kw for interrater agreement for the items “hold the animal” (0.32±0.24) and “return to the cage” (0.27±0.24) were low, despite the fact that the same items presented higher percentages of agreement (> 95 %). For the remaining scale items, including the “global impression” item, there was moderate interrater agreement (0.48 ±0.11 ≤ Kw ≥ 0.56 ±0.12) with percentages ranging from 76 to 88%. The first scale item “remove the animal from cage” showed good coefficient (Kw = 0.75±0.12) and percentages of agreement (97%) (Table 2 – B).

Regarding the mouse intraperitoneal administration scale, it was not statistically possible to calculate kappa values for restraint and injection speed items. The percentages of agreement were acceptable (≥ 97%). One item in this scale, “body part quadrant”, revealed low interrater agreement (Kw= 0.33±0.15) and high percentage of agreement (90%). The remaining scale items presented moderate (“depth of puncture/angle” and “global impression”), good (“animal position” and “reflux/syringe holding”) and excellent (“animal position” and “time after restraint”) kappa values and the percentages of agreement ranged between 75 and 100 (Table 2 – C).

For some items of the rat subcutaneous administration scale (“body part/quadrant”, “injection speed”, “time after restraint” and “procedure repetitions”), similarly to intraperitoneal administration, it was not possible to calculate Kappa due to statistical assumptions behind Kw calculation. Nevertheless, those same items showed acceptable percentages of agreement (96 to 100). The kappa of the remaining items was moderate (“depth of puncture/angle” and “global impression”) and good (“restraint/animal position” and “reflux/syringe holding”). Percentages of agreement for these scale items ranged between 83 and 95 (Table 2 – D).

The "global impression" (GI) item of each scale presented moderate values of kappa and acceptable level of percentage of agreement: GI for mouse restraint scale Kw=0.51±0.07 and 71%, GI for rat restraint scale Kw=0.49±0.10 and 76%, GI for mouse intraperitoneal scale Kw=0.57±0.09 and 75% and GI for rat subcutaneous administration scale Kw=0.59±0.11 and 83% (Table 2).

Table 2 | Calculation of Weighted kappa (Kw) and percentage of agreement (% agreement) between raters for all items of the four different GRSs for the total of performances (n performances); (*) One of the variables is constant therefore Kw cannot be computed (a) Each variable in each pair has less than two valid categories therefore Kw cannot be computed.

A

MOUSE RESTRAINT					
Scale item		n performances	% agreement	Kw	SD
pick from the cage	tail grasp attempts	92	96	-.02	.01
	tail hold	92	85	0.18	0.12
	suspension duration	89	96	0.31	0.25
restraint & scruffing	neck scruff attempts	89	78	0.54	0.09
	scruffing	92	76	0.51	0.09
	tail holding/restraint	92	85	0.45	0.13
after restraint	procedure repetition	92	75	0.61	0.08
global impression		92	**71**	**0.51**	0.07

B

RAT RESTRAINT					
Scale item		n performances	% agreement	Kw	SD
pick from the cage	remove the cage	88	97	0.75	0.12
	hold the animal	88	95	0.32	0.24
restraint & scruffing	cranial/anterior restraint	88	82	0.48	0.11
	caudal/posterior restraint	88	88	0.56	0.12
after restraint	return to the cage	88	97	0.27	0.24
	procedure repetition	88	78	0.55	0.10
global impression		87	**76**	**0.49**	0.10

C

MOUSE INTRAPERITONEAL ADMINISTRATION					
Scale item		n performances	% agreement	Kw	SD
before admin.	restraint	88	97	*	*
	animal position	88	92	0.90	0.04
administration	body part (quadrant)	88	90	0.33	0.15
	depth of puncture /angl	87	75	0.54	0.08
	reflux /syringe holding	88	80	0.74	0.09
	injection speed	88	99	*	*
after admin	time after restraint	88	100	1	0
	procedure repetition	88	99	0.79	0.20
global impression		88	**75**	**0.57**	0.09

D

RAT SUBCUTENEOUS ADMINISTRATION					
Scale item		n performances	% agreement	Kw	SD
before admin.	restraint/animal position	84	95	0.82	0.9
	body part - quadrant	84	98	*	*
administration	depth of puncture/angle	84	89	0.59	0.13
	reflux/syringe holding	84	89	0.71	0.10
	injection speed	82	100	*	*
	time after restraint	83	96	*	*
after admin	procedures repetitions	84	100	error	
global impression		84	**83**	**0.59**	0.11

The OSLASE acceptability for trainees

From the focus group interviews performed with former trainees, five topic categories emerged related with the practical exam as illustrated in Figure 1. Verbatim quotes are given as examples to clarify and confirm results. Respondent pseudonym and course format are indicated at the end of each quote.

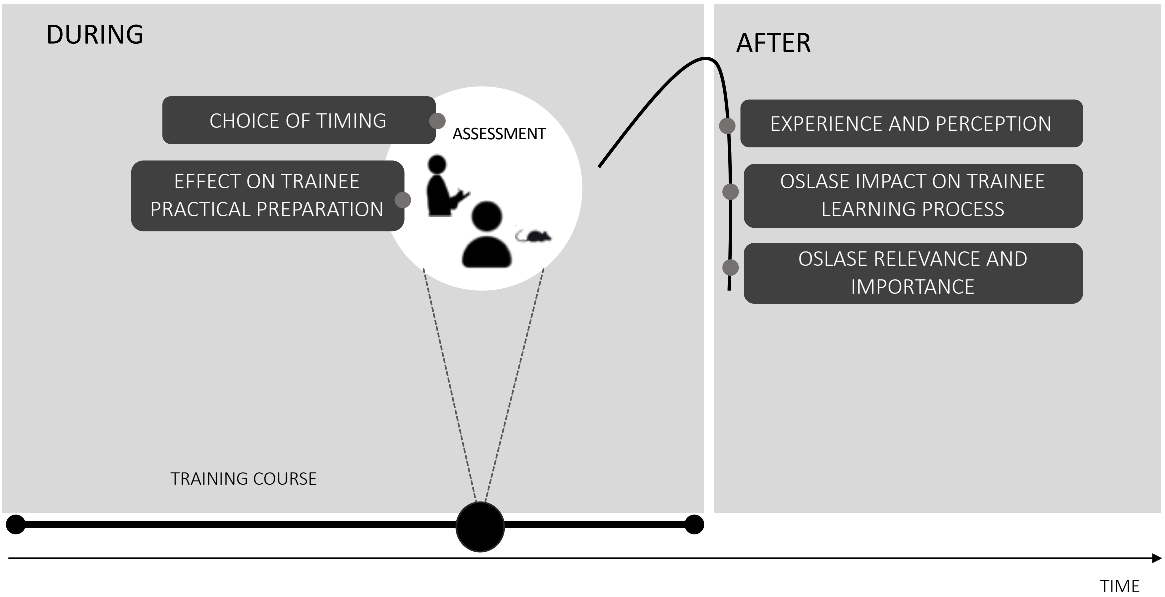

Figure 1 | Illustration of focus group categories and their relation with the competence assessment timepoint at the course. Five categories emerged: during the course (1) *Choice of timing* to perform the exam (2) and *Effect on trainee practical preparation* before the OSLASE, and after the assessment (3) *Experience and perception*, (4) *OSLASE Impact on trainee learning process* and (5) *OSLASE relevance and importance*

(1) *Choice of timing*

Course participants were asked about the time chosen to perform the OSLASE. For interviewees from the advanced course, feeling confident and making the best use of the available time to practice before the exam were key considerations in the choice of when to do the demonstration of practical skills in rodent handling.

Only one of the interviewees reported that they chose to do the demonstration at the first possible opportunity:

> *I asked to [do the demonstration] during the first [class] because as I said, I was already more at ease with handling animals (…) I had done the other, more basic, course and it had the same practical component, therefore I felt comfortable to do it in the first class.* (SB, advanced)

The remaining interviewees had chosen the second animal handling class (end of handling module) for their practical demonstration, for reasons that were all variations on the need to practice more before they felt confident:

> *I, who had no experience in handling animals, felt that in the end of the first class was too early. It might even go well, let's say, but I don't know if I would have performed in the best possible way. Therefore, I thought it would be safer, I would feel safer doing it in the next class, given that this was a possibility. And then at the end of the second class I was more comfortable, I was more used to the technique and I think then it was easier too.* (RCP, advanced)

Without exception, interviewees from the introductory course reported having chosen to do the practical exam after they had concluded the training period with their tutor, rather than directly after the end of the practical classes. These students, whose practical training had been impacted by restrictions during the Covid-19 pandemic, also highlighted the balance between avoiding a too long gap but still having enough time to practise: "*a month for me would be a little too much between the practical classes and the exam*" (CP, introductory).

(2) *Effect on trainee practical preparation before performing OSLASE*

Several of the interviewees from the advanced course expressed that the existence of a formal practical exam enhanced their engagement in the practical class of animal handling:

> *It makes things more formal as regards the learning. I think we engage more in wanting to do it and to do it well because we know we will have to demonstrate later on this same knowledge that we learned* (JA, advanced)

Others, however, argued that they were not approaching the learning any differently as a consequence of knowing they would have to pass an exam:

> *I don't think it had any effect at all, in any case I was there to learn and certify that I had learned to work with mice, and therefore the demonstration was a validation in addition to my own, only that* (CA, advanced)

Among interviewees from the introductory course, there was also a widespread view that knowing there would be a formal practical exam had little influence on how students prepared and practised. One student explained that this was because they were *"expecting there to be a practical exam because this makes sense for this course and therefore I didn't prepare differently, I prepared according to what I was expecting".* (MN, introductory)

Other students referred that because they were preparing for real life work with live animals, their preparation was not dependent on there being an exam:

> *For me no [it didn't affect how I prepared] because even some of the procedures were simulated in toys, not in live animals, I knew afterwards I will need to do them in live animals. So for me, I knew I had to do the procedures well because then I*

was going to do them in live animals anyway, so for me it didn't make a difference. (LM, introductory)

(3) *Experience and perception*

Some respondents reported that they unavoidably felt anxious at having to make a practical demonstration that implied "*going into another room and saying it is an assessment, even if one's a pro in that, it's always a moment of some nervousness*" (SB, advanced). Respondents reporting that they did not feel much pressure referred to "*the preparation we had before allowed this demonstration to be made in a tranquil way*" (RPC, advanced) and the course having "*few students [and] in turn the relation to the instructors was easy*" (IL, introductory).

(4) *OSLASE Impact on trainee learning process*

Among interviewees from the advanced course, where the formal evaluation of handling skills takes place during rather than at the end of the course, it was perceived to have a positive impact on subsequent course moments by providing "*an injection of confidence in animal handling"*, perceived to be "*very important (…) before the more advanced procedures* (PA, advanced). External evaluation was found to provide "*more confidence in ourselves and we have that internal dialogue that a competent person determined us as evaluated and that we perform handling well, and therefore we can proceed with our work*" (JC, advanced), and contributing to *"completing the rest of the course with greater commitment"* (CA, advanced).

An evaluation between the first and the subsequent classes may also contribute to levelling the standard between students:

> *Perhaps if this evaluation would not be there, the first classes would be habituation but probably in the next classes there may be members who would not feel so at ease with the activities, feel some difficulty, hence these two classes with evaluation ensure that we are more prepared for the subsequent classes which are more difficult in terms of procedures.* (CA, advanced)

Regarding whether the exam changed students' perception, some thought "*that the practical assessment in itself made no difference (…) it was really the theoretical assessment*" (CP, advanced), whereas others reported that "*in terms of experience and really dealing with the animals, handling them and doing the necessary protocols is really different with a theoretical assessment and a practical assessment*" (MN, introductory).

(5) *OSLASE relevance and importance*

Interviewees considered the *OSLASE* was an essential component of the course because "*if this is a course [that aims] to prepare us much for practising with laboratory animals the practical assessment cannot be missing*" (MN, introductory), whereas others expressed a mixed view: "*if there were no evaluation would people learn? Yes. because I think all the people who were there were diligent, but taking into account the course it makes sense to have a [practical] exam*" (SB, advanced)

One of the participants highlighted the relevance from the perspective of responsibility for overseeing work with animals in that not having a practical exam *would place a much greater responsibility on the supervisor of the practical procedures right after the exam* (IL, introductory).

Trainees in the introductory course considered the format was adequate, as they "*were assessed in the most basic, but also in the most common, animal manipulation and restraint, and the most common injections*" (ILG Intro), whereas for the advanced course "*it was at the right moment and represented what we had learned before* (CA Advanced) and "*without knowing [how to handle animals well] it would be difficult to proceed to the rest" (JC Advanced).*

Regarding the specific relevance for the respondent in question, one found that "*only after the assessment did I effectively understand that there were some specific points that need improvement*" (MN), whereas several respondents said that it was not so much the exam itself but the training that helped them see where they needed to improve.

Interestingly, one of the respondents discussed the relevance of a practical exam in the context of having "*another tool for talking with people about the topic and be able to talk about one's own experience regarding the level of preparation and the level of training*". This allows them to respond to colleagues from other fields claiming that ethics of animal experimentation are empty words that do not translate into practice, by saying "*look it's not quite that way; to do all this, we passed through training, we passed through verification*" (IL Intro).

Discussion

The aim of this study was to explore the implementation of OSLASE in laboratory animal science training, evaluating the interrater agreement in the practical assessment and the trainees´ acceptance of the exam.

The consistency, or the agreement among raters provides valuable information regarding the quality of assessment and its reliability. Particularly, it reflects if raters agree on how to apply the instrument as a support for deciding on a candidate's performance. Two statistical methods that have been used in health sciences (13, 19) were used to calculate Interrater reliability (IRR): percentage of agreement and weighted Kappa (Kw). Both statistical measures of agreement present advantages and limitations: percentage of agreement is easily calculated and directly interpretable, but does not take into account whether the raters are scoring randomly. Kw formula corrects this limitation but is based on assumptions regarding rater independence and other factors which are rarely met in real life data, which may result in extreme underestimation of the level of agreement (13). For this study results, the majority of the items of all the four scales revealed moderate to good values of Kappa as well as acceptable percentages of agreement for all items, except one. The Kw and agreement percentage showed similar results, confirming a tendency on agreement, although some scales items are not fully concordant on the magnitude of accordance. Previous studies have reported moderate kappa values and excellent percent agreement between human and automated raters (20).

Based on these results, there is room to enhance IRR in the application of the OSLASE. One way of doing so is by training the assessors further in how to apply the scale. Continuous rater training is challenging to implement into busy teaching schedules, but it has been shown to considerably minimize disagreement, particularly when scale items are measuring fine and very technical discriminations with several rating categories (as the scales used in this study) (21, 22). The performances included in this study were assessed over a 3-year period, and despite being administered by assessors that are very familiar with the GRSs, the quality of the data collected was not monitored over time to find potential disagreements and this fact may have contributed to the limited IRR scores. Also, scales were applied to a restrict number of performances in a controlled scenario (of training). Diversifying the contexts in which the instrument is applied, such as by including workplace assessments (of researchers performing procedures as part of their work with animals) as a part of a continuing professional development strategy could enhance IRR (23, 24).

Going beyond the question of agreement, it is also relevant to consider the value of the OSLASE as a way to implement an assessment strategy able to evaluate the practical competence acquired by researchers during training (for advanced course format) or at the training end (for introductory format). This adapted OSCE type exam introduced standardization and uniformization of the assessments for all trainees (independently of the course type). The existence of standardized instruments introduces predictability for the trainees, who know that and how they will be evaluated, and can get systematic feedback on their performance. Instruments applied at the stations, the GRSs, are promoting consistency which guarantees that how the candidates' performance is evaluated is not dependent on who is performing the assessment.

Regarding OSLASE acceptability by students, findings reflect that trainees accept and recognize the validity of this type of structured practical assessment on the context of

LAS training course. Trainees took into consideration their confidence in their own capacity to decide on the timing to take the OSLASE and this seems to be the key factor affecting the decision of when to perform the exam. It is noteworthy that trainees also highlighted another time issue: the time that is needed to develop and reach a level of proficiency to succeed in the assessment. Trainees who were forced to postpone the OSLASE due to the pandemic restrictions reported frustration with the delay of the assessment, recognizing that the exam should be adjusted to the training timeline, as an excessive time interval between the end of classes and the exam was perceived as problematic by trainees. The integration of OSLASE in the course program affected trainees' preparation: some respondents recognized that they felt more engaged and stimulated to invest on practical training. Furthermore, trainees pointed out that assessment ensures that apprentices reach the expected practical learning outcomes. These findings are in line with research on the topic that supports that assessment not only drives but also has a positive impact on learning (6, 8). Whereas the OSLASE is a relatively recent introduction in this training, a few respondents stated that for them a formal practical assessment was assumed and expected in these courses. Assessment was not a stressful experience for the majority of the respondents, and this fact may be justified by the way the experienced tutors, also in the role of assessors, drove the experience. Indeed, the assessor role being assumed by the tutors can be faced as a limitation of the study which can bias the way candidates faced the OSLASE. However, the LAS tutors with expertise to teach, supervise and assess are limited in number. Interviewees also identified the positive impact of the exam on the learning process, as previously reported in the literature (9). The topic of how OSLASE helps the trainee to confirm the skills acquired emerged several times during interviews, confirming the perception that assessment is an essential step to consolidate and promote future practical learning (25, 26). Trainees also reported that OSLASE establishes a standard and defines which skills students are expected to develop in the next sessions. They also felt that demonstrations contributed: to identify skills that should be improved and to support the role of those overseeing workplace training, which are crucial features that assessment introduces in the context of these courses.

Conclusion

To the best of our knowledge, this is the first study reporting the impact of implementing a practical assessment strategy in a laboratory animal science training context. In particular, this study addressed the feasibility of OSLASE by exploring interrater agreement and investigating trainees' acceptance of the practical exam. The results show that there is room for improvement in interrater agreement, but overall raters agreed on the decisions regarding trainees' competence. The trainees recognized that

OSLASE promoted and guided their learning, increased their engagement and motivation during practical sessions, as well as served as a confirmation that relevant skills had been learned. Overall, the introduction of OSLASE allows the measurement of practical competence besides the cognitive knowledge assessed with traditional evaluation methods.

Acknowledgements

The authors would like to give a special acknowledgement to all course participants who collaborated in this study. The authors also thank Dr Patricio Costa, from the University of Minho, for statistical assistance.

Ethical Approval

Data collection with human participants (instrument application in the courses context and group interviews) were performed under approvals from the Joint Ethics Committee CHUP/ICBAS with the reference 2019/CE/P014(292/CETI/ICBAS) and 2021/CE/P013(P353/CETI/ICBAS).

Informed consent was obtained by candidates prior to data collection.

Data availability statement

Data for this study were collected from human participants. They consented for the use of data for research and publication purposes, but not for providing as open data. Raw data collected for Interrater Reliability calculation will be available upon request to corresponding author.

References

1. European Commission guidance document. National Competent Authorities for the implementation of Directive 2010/63/EU on the protection of animals used for scientific purposes. A working document on the development of a common education and training framework to fulfil the requirements under the Directive. Brussels, http://ec.europa.eu/environment/chemicals/lab_animals/pdf/guidance/education_training/en.pdf (2014, accessed 13 April 2020).
2. European Parliament and Council of the European Commission Union. Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. Strasbourg, France: Council of Europe, 2010.
3. NC3Rs (National Centre for the Replacement RRoAiR. The 3Rs 2022 [Available from: https://nc3rs.org.uk/the-3rs.

4. Duffee N, Nevalainen T, Hau J. Education and Training. In: Hau J, & Schapiro SJ, editors. Handbook of Laboratory Animal Science: Essential Principles and Practices 4th ed: CRC Press; 2021. p. 12.
5. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990; 65(9Suppl):S63-7.
6. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. The Lancet. 2001;357(9260):945-9.
7. Bashook PG. Best practices for assessing competence and performance of the behavioral health workforce. Adm Policy Ment Health. 2005;32(5-6):563-92.
8. Tabish SA. Assessment methods in medical education. Int J Health Sci (Qassim). 2008;2(2):3-7.
9. Preston R, Gratani M, Owens K, et al. Exploring the Impact of Assessment on Medical Students' Learning. Assessment & Evaluation in Higher Education. 2020;45(1):109-24.
10. Costa A, Lamas S, Costa MJ, et al. The assessment of researchers' competence in experimental procedures with laboratory animals: A three-step methodology to develop a global rating scale. Lab anim. 2021;55(5):463-71.
11. Costa A, Lamas S, Costa MJ, et al. Global rating scales to assess the competence of researchers in handling and performing substance administrations in laboratory rodents [Internet]. 2021 [cited 19/04/2021]. Available from: https://repositorio-aberto.up.pt/handle/10216/133419.
12. Harden RM, Stevenson M, Downie WW, et al. Assessment of clinical competence using objective structured examination. *Br Med J* 1975;1(5955):447-51.
13. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-82.
14. Graham MJ, Milanowski AT, Miller JB, editors. Measuring and promoting inter-rater agreement of teacher and principal performance ratings [Internet]. 2012 [cited 19/04/2022]. Available from: https://files.eric.ed.gov/fulltext/ED532068.pdf.
15. Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of kappa and weighted kappa. Psychological Bulletin. 1969;72(5):323-7.
16. Cohen J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological Bulletin. 1968;70(4):213-20.
17. Chmura Kraemer H, Periyakoil VS, Noda A. Kappa coefficients in medical research. Stat Med. 2002;21(14):2109-29.
18. Altman DG, editor. Practical statistics for medical research. London: Chapman and Hall; 1991.

19. Hripcsak G, Heitjan DF. Measuring agreement in medical informatics reliability studies. Journal of Biomedical Informatics. 2002;35(2):99-110.
20. Simundic AM, Nikolac N, Ivankovic V, et al. Comparison of visual vs. automated detection of lipemic, icteric and hemolyzed specimens: can we rely on a human eye? Clin Chem Lab Med. 2009;47(11):1361-5.
21. Kardong-Edgren S, Oermann MH, Rizzolo MA, et al. Establishing Inter- and Intrarater Reliability for High-Stakes Testing Using Simulation. Nurs Educ Perspect. 2017;38(2):63-8.
22. Holland AE, Tiffany J, Blazovich L, et al. The effect of evaluator training on inter-and intrarater reliability in high-stakes assessment in simulation. Nursing Education Perspectives. 2020;41(4):222-8.
23. Crossley J, Humphris G, Jolly B. Assessing health professionals. Med Educ. 2002;36(9):800-4.
24. Van Dalen J, Prince CJAH, Scherpbier AJJA et al. Evaluating Communication Skills. Adv Health Sci Educ Theory Pract. 1998;3(3):187-195.
25. Bidonde J, Meneses-Echavez JF, Asare B, et al. Developing a tool to assess the skills to perform a health technology assessment. BMC Medical Research Methodology. 2022;22(1):78.
26. Heeneman S, Oudkerk Pool A, Schuwirth LW, et al. The impact of programmatic assessment on student learning: theory versus practice. Medical education. 2015;49(5):487-98.

7. GENERAL DISCUSSION

GENERAL DISCUSSION

Providing an adequate training in Laboratory Animal Science (LAS) and ensuring that the knowledge and practical skills have been acquired by the researchers are crucial to guarantee the welfare of laboratory animals used in experiments. In Europe, educational standards to conduct experiments with animals have been established by governmental and non-governmental organizations (1-3).

To the best of our knowledge, there has been no systematic research into the impact and acceptance of different training approaches and the e-learning use in LAS field and the work published regarding the use of e-learning in LAS was limited to a few congress presentations (4, 5). Preliminary research on the topic of proficiency assessment has been reported by Clifford et al. (2013). Their work highlighted the relevance of competence assessment in training programs and in standards improvement, ensuring consistent training, while minimizing errors and guaranteeing animal welfare. The authors highlighted the important role of score sheets to guide training and to assess competency, in a standard and objective way (6).Other authors have reported strategies for the assessment of competence in LAS courses, describing their experiences based on the administration of different types of assessments (7).The Directive 2010/63/EU adoption, requiring staff assessment and competence demonstration, together with the implementation of Continuing Professional Development programs, lead to the dissemination and adoption of the first instruments to assess researchers' techniques, procedures and practical skills. These evaluation instruments are in the format of Direct Observation of Practical/Procedural Skills (DOPS). Laboratory Animal Science Association (LASA) website shared the first tranche of DOPS developed for Royal Veterinary College(8). Also, recently, in 2022, the Education and Training Platform for Laboratory Animal Science (ETPLAS) published on its website the first set of DOPS, to be used as examples, for rodents and zebrafish. This work was developed by a working group chaired by Lucy Whitfield and five members including the thesis author, Eirini Fragiadaki, Rafael Frias, Brian Mphande and Paul West, as part of the ETPLAS Pilot Project "Promoting alternatives to animal testing through accessible and harmonised education & training" intended to create a database of common practical tasks assessment (9). Despite the dissemination of these first assessment instruments and the efforts to standardize assessments in LAS, there is still a lack of research on the topic. This is in contrast with medical and to some extent veterinary education, where research has provided essential understanding of how to develop instruments, implement strategies, and assess their impact.

This thesis focus were the two probably most prominent topics on the training and education in Laboratory Animal Science: delivery of content and assessment of competence.

7.1 | TRAINEES' PERCEPTION OF E-LEARNING

The first study presented in this thesis (chapter 4) reported the general acceptance and satisfaction of LAS trainees after the adoption of e-learning using the Moodle platform implementation for the theoretical component delivery (as a combined or full approach).

The assessment of satisfaction and acceptance requires appropriate instruments, which should meet standards for psychometric validity and be adjusted to the context of administration when developed (10). Nevertheless, in the LAS educational field, there is a lack of instruments to measure e-learning perception, and the existing were not adjusted to this particular context (11-13): intensive postgraduate training providing biomedical researchers the knowledge and skills to perform experimental procedures with laboratory animals. Therefore, this first study initiated a precursory effort to develop a measurement instrument, the Questionnaire of E-learning Acceptance (QELA), adjusted to LAS context, able to explore and measure the trainees' perceptions regarding e-learning approach. The questionnaire was created considering the existing instruments for e-learning evaluation and acceptance developed in other research fields. The instrument design aimed to explore LAS trainees' perception in several domains, previously identified to cover relevant and crucial features of e-learning approach (11-13).Thus, QELA comprises five subscales: a) Personal perception of the e-learning use; b) Satisfaction with the e-learning content; c) Satisfaction with the e-learning organization; d) Time management and e) E-learning practical influence. This last scale was developed in order to clarify the potential impact and influence of e-learning use in the LAS practical sessions.

Independently of the strategy used to combine e-learning approach with classroom lectures, which was different in introductory and advanced training courses, high levels of acceptance of e-learning were reported. Particularly, this was found in the domains of Personal perception of e-learning use, Satisfaction with content, and Satisfaction with platform organization. These results were in line with findings in the veterinary field, with students reporting positive perception of e-learning (13-15). Our results also confirmed that LAS trainees recognize the time management efficacy and time control that e-learning allows and the advantage introduced by this approach in these intensive course formats. Previous research in veterinary sciences assessing the use and acceptance of e-learning and its impact on distance learning, particularly regarding time control, corroborates these results (16).

Another aspect that emerged in this study was the perception by the overwhelming majority of respondents about the positive influence that e-learning may have when trainees prepare for practical training. For many trainees, their first contact with rodents was in the practical sessions, which could be a stressful experience for students coping with several new aspects for the first time. It is noteworthy that in this study 32.6% of the total respondents indicated that they had no previous experience with rodents, and that 16.7% indicated fear of working with laboratory animals. Nevertheless, QELA administration revealed that e-learning use in LAS theoretical training seems to have beneficial effects on the promotion of the practical learning, as well as in levelling this experience of first contact with rodents. One possible explanation for these results might be that the online availability of lectures allowed trainees the opportunity to revise animal handling and practical techniques, as many times as they feel necessary and at their own pace. Despite these assumptions of potential impact of e-learning on the promotion of trainees first contact with rodents, when directly questioned, study respondents did not recognize that "previous experience" and "fear" has affected their perception of e-learning use regarding practical classes.

Indeed, acceptance of the e-learning seems to be highly influenced by previous experience with laboratory animals, with inexperienced students from both groups (advanced and introductory) showing higher levels of platform acceptance. The advanced sample, with a heterogeneous background regarding respondents' previous experience with laboratory animals, showed that more experienced trainees were less satisfied with e-learning organization and content, with lower perceived acceptability of e-learning. Educational background was identified as a factor also affecting e-learning acceptance. The influence of demographic factors in e-learning acceptance and use has been reported in other studies (17, 18). Also, the health sciences educational literature suggests students' levels of motivation (internal and external) and expectations impacted levels of e-learning acceptance (19).

There are limitations to this study. Despite the large sample size (229 respondents) for a preliminary study in this context, findings' interpretations and generalization require some caution, particularly due to limited number of participants for some categories (such as background education), which may not be representative. LAS courses tend to have very heterogeneous participants regarding background and previous experience with animals, which in terms of design would demand a larger sample size in each group. Therefore, several other factors may be influenced by the respondents' perception, such as the research context and the need of animal use, the training urgency or even other factors such as prior online learning experiences. The impact of those factors should be explored in future studies, and a larger study would be desirable to confirm these preliminary results.

The distribution of total acceptance variable shows a negative skewness effect (also known as left-skewed) created because respondents tend to answered positively to the scale items. This may suggest that QELA, the instrument used, may need further refinements to better reflect and discriminate the less accepted features of e-learning approach.

In spite of these limitations, this work provides the first detailed analysis on LAS trainees' views, satisfaction and acceptance with e-learning reporting relevant clues, namely regarding features affecting its acceptance, which might be considered to structure and design courses using e-leaning. Our results are supported by previous studies in veterinary educational context, where e-learning could not fully replace particular teaching approaches, but promoted the learning outcomes attained, which are difficult to reach with traditional teaching methods (14, 15).

The introduction of e-learning in our LAS courses was motivated by the organizers´ increasing concerns regarding the extensive classroom program and curricula. The course intensity and the lack of program flexibility had often been reported as negative aspects by the course participants in the feedback forms. Particularly in the introductory courses, there was also an institutional interest in developing a training solution for frequent delivery of high-quality training for new staff needing training to work with animals. We started to integrate e-learning in the courses to complement theoretical sessions around 2010.The questionnaires to explore trainees' acceptance were administrated between 2012 to 2015, and we submitted this work (chapter 4) for publication in 2019. The emergence of COVID-19, in 2020, and the need to rapidly provide alternatives to classroom teaching brought other perspective and relevance to the findings published in this study. With the outbreak of COVID-19, teachers in LAS as well as in other educational areas, were forced to implement adjustments in mandatory training, and the interest in e-learning grew dramatically. Our study provided important insights into e-learning as a complementary approach to support theoretical LAS training, as the results confirm that e-learning is accepted by trainees in this context. This work is being cited in studies exploring the impact on health education during pandemic (20-22). At our institution, during this period, theoretical LAS training was maintained with minor adjustments. Students were invited to complete online modules according a pre-established program combined with synchronous online classes, during the COVID-19 lockdown.

7.2 | ASSESSEMENT OF COMPETENCE

Assessment of knowledge and skills is an essential part of the education and training process. Nevertheless, much can be assimilated and integrated from education areas with contextually similar challenges, such is the case of human medicine, where there has been much more research.

In the scope of this thesis and under this topic, we started by proposing a methodology to develop an assessment instrument (chapter 5) able to capture the higher levels of Miller's pyramid of assessment (23). An instrument applicable to performative and behavioural dimensions of performance, whether in controlled/simulated or in real/ professional settings. We considered four crucial and representative stations, assessing basic skills to perform with rodents.

The marking of candidates' performances in observational assessments can be supported by instruments of different types, the most usual being Global Rating Scales (GRSs) and Checklists (CLs). We decided to develop GRSs, which enable the distinction between different levels of proficiency and expertise, and dimensions not represented in CLs (such as time, efficacy perception and safety). According to existing iterates from other educational fields, GRSs allow for greater discrimination between performances, while CLs have lower sensibility and the same item executed with different proficiencies can be scored with "yes" (24-26). Consequently, GRSs provide more detailed information for trainees and assessors. Therefore, in LAS, GRSs seemed to be the most appropriate instrument to support formative (assessment for learning) and summative assessments (assessment of learning) in different scenarios.

The methodology to develop GRSs proposed in this work is not limited to GRSs and can guide the development of others scale's types, CLs or DOPs, for different species and procedures.

The establishment of a core team that supervised and integrated the feedback received during different steps of the instrument development was essential to ensure a dynamic iterative process and the successful conclusion of the scales' development. The expertise of the different core team members was decisive for the process, namely in the definition of skills to be assessed and in the development of instruments. The core team members with experience in LAS training understand the most critical aspects of the procedures that can compromise animals' welfare and interfere with the performance' success. On the other hand, they also have the knowledge to decide which aspects of raters' feedback should be integrated in the scale items and how. Members with expertise in implementation of performance assessment instruments provide crucial knowledge regarding psychometric properties, essential to initiate and to guide the creation of a scale.

In this work, the two assessors which administrated and tested the scales' applicability were also integrated in the core team. This circumstance can be considered as an advantage regarding the feedback input, but can also be regarded as a limitation, since a bias approach may condition the potential identification of the scale aspects that should be improved or amended. To overcome this possible bias, we limited the number of assessors integrating the core team to two persons.

We have used videos in several stages of the scale development. The possibility to support the scales development and the raters training with dedicated videos is particularly advantageous in this area. Despite the potential limitation of the camera angles or image quality, which may limit the performance observation, this resource can effectively bring great improvements to the scales and the raters' objectivism training. They allow representing different types of performances and can be used repeatedly, without using living animals. In the workshop organized at FELASA (appendix IV) videos performed in mannequins, were used at the beginning of the session, to promote a first approach of the group to the procedure and initiate the creation of scales. At this stage, simple videos were useful to trigger the discussion of which skill items descriptors should be integrated in the scale. In a process of scale refinement or implementation, videos collected from several sources (YouTube, educational websites, etc) revealed to be useful to establish correct and adjusted descriptors and to train assessors to apply the instrument. Based in our experience, the use of videos allows to identify and bring into discussion potential problems and ambiguous situations that may occur in authentic scenarios. In fact, for scale refinement and assessors training purposes, videos should as diverse as possible.

After conclusion of scales' development, to be used to assess performance at manned/observed stations, we move on to the next step, the implementation of the OSLASE - Objective Structured Laboratory Animal Science examination, based on the original OSCE – Objective Structured Clinical Examination described by Harden (27).The OSLASE's impact, particularly the exam reliability and its acceptance by the trainees, was explored during the researchers' training (chapter 6). In the context of our studies, the OSLASE, comprising four exam stations, was integrated at the end of introductory training as a final assessment, with the objective of complementing the candidate's score obtained with the traditional written assessment. In the advanced courses, the exam was adjusted to focus only in the handling and restraint competencies, therefore including only two stations for both species (mouse and rat). The OSLASE was implemented at the end of the handling module, which is the first of the practical modules integrated on the advanced course program. Candidates may choose the day to perform the OSLASE in the two course formats.

The trainees' acceptability regarding the OSLASE, in both course editions, was explored using focus group interviews, a qualitative approach method. The interviews findings revealed that trainees proposed themselves for OSLASE based on their feeling of confidence in being able to perform the procedure and, therefore, the "time" for exam preparation was highly valued by them. Reflections by students who had experienced the impact of Covid-19 restrictions which lead to suspension of introductory training, highlighted that LAS assessment should be synchronized with training, and giving more time than necessary for practice was not appreciated by trainees. Similarly to what has been reported in the literature for other contexts, LAS trainees recognized that the existence of a practical assessment increased their motivation and engagement in the training (28, 29) and had a positive impact on their learning (30). According to trainees, assessment is helpful to identify procedures' aspects that they need to improve in future training sessions/period of supervised work. Also, passing the exam confirms to the candidate that the skills expected were acquired, something that was important to consolidate their confidence. The finding that assessment consolidate and promote future practical learning is in line with previous research (31, 32)

To explore interrater reliability (IRR), weighted kappa and percentages of agreement were calculated for two assessors which applied the developed scales (chapter 5 – Appendices II) within 92 performances. Weighted kappa (Kw) revealed a moderate to good IRR for the majority of the scales' items as well as acceptable percentages of agreement. Information collected by assessors with scales are producing acceptable judgements of trainees' performance. Nevertheless, agreement must be improved with training. This training should be faced as a continuum requirement, it is essential not only before the scale's administrations but also at different timepoints if scales are being administered systematically.

Regarding scales applicability, for the two study' assessors they revealed to be clear, of easy rating and, effectively, able of identify performance' items and categorize the observed skills during trainee evaluation. Indeed, when scales are used in formative training contexts independently by assessors' observing the same performance, the feedback produced tend to focus and identify the same aspects that should be improved by the trainee.

The global impression item, the final item of all scales, reflects the assessor's overall impression of a candidate's performance. Indeed, it captures aspects that were not directly measured at the scale (such as professionalism, empathy with animals, etc.) which provides crucial information for the assessor final decision.

In both courses, candidates had to achieve the standard defined for the OSLASE to be approved, and in case of failure (performance did not reach the minimum established for the station) the candidates had to retake training.

The standard established for the OSLASE was adjusted to the training context and adapted to the tutors' expectations for those skills that newly trained researchers must have. When setting the standard for an assessment it is important to consider the purpose of the assessment and the motive to establish the cut score (33-35). Briefly, key aspects that must be taken in account include knowing the assessment content and the candidates' characteristics; establishing what is expected from candidates when authorized to perform unsupervised work and/or candidate future responsibilities and/or functions that training must ensure. All of these should be considered prior to practical assessment, independent of the type of assessment.

From the courses organizational point of view, the OSLASE implementation, particularly in the advanced course introduced a great advantage in the dynamic of practical handling sessions. The OSLASE strategy in these sessions, where students showing that they have acquired the necessary competence are signed off, allowed that only trainees which need to develop practical skills stayed in the class. The tutors were able to dedicate more time to each trainee who effectively needs support. A calmer environment with fewer student positively impacts students' learning, and is highly recommended when handling living animals.

The assessment studies presented in this thesis report findings limited to a particular institution and context, with a sample size of both trainees and assessors that may be considered small. Undoubtedly, it is essential that further research explores the implementation of OSLASE and other types of practical assessment in different LAS environments and with a more heterogeneous group of trainees. This will be a crucial step to gather knowledge and understanding that will allow establishing and ensuring standards around different institutions in training and working with lab animals.

It is important to highlight that research on education clarifies that in order to infer the candidate proficiency based on valid and reliable information, the performance evaluation requires using a number of different assessment methods, considering the distinct methods' limitations and strengths. In LAS training it would be desirable not only to assess candidate's proficiency in a single circumstance, such as end of training period, but also throughout several timepoints, whenever considered relevant for the practical work and for the procedures using animals planned at the study. Also, using different assessment methods, such as DOPs along with GRSs or CLs, would be recommended. Every skill has a "learning curve" followed by a "forgetting curve", therefore, testing a candidate only at a specific time point of the learning process does not provide information on whether or not the trainee will retain the skill (36, 37).

As pointed out in the introduction of this thesis, educational research is still exploring and clarifying the relation between competence and performance: how can the first predict the second, or how performance is observed while competence is inferred (38, 39). It is important to consider that in LAS, similarly to other areas, the challenge remains and the information and evidence that is possible to capture when observing a candidate demonstration in a high-stake assessment is limited, even using valid, reliable and adjusted measurement instruments. Thus, to the best of our knowledge, to ensure that researchers are competent and properly prepared to work with living animals and to establish standards in laboratory animal science, more assessment strategies must be adopted.

REFERENCES

1. European Commission guidance document. National Competent Authorities for the implementation of Directive 2010/63/EU on the protection of animals used for scientific purposes. A working document on the development of a common education and training framework to fulfil the requirements under the Directive. Brussels, http://ec.europa.eu/environment/chemicals/lab_animals/pdf/guidance/education_training/en.pdf (2014, accessed 13 April 2020).
2. European Union. 2010. Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. Official Journal of the European Union L276/33.
3. FELASA. Education & Training Accreditation Scheme. Available from: https://felasa.eu/education-training/course-listings (2021, accessed 13 April 2022)
4. Exner C SN, Heldmaier G. E-learning, A New Training Tool for the Qualification of Students in the Handling of Laboratory Animals. FELASA-ICLAS Joint Meeting; 2007 11-14 June; Villa Erba,Cernobbio (CO) Italy.
5. Linklater NEC, Heldmaier G. eLearning in LAS to meet education and training requirements. 12th FELASA SECAL Congress; 2013 10-13 June; Barcelona.
6. Clifford P, Melfi N, Bogdanske J, et al. Assessment of proficiency and competency in laboratory animal biomethodologies. J Am Assoc Lab Anim Sci. 2013;52(6):711-6.
7. Hansen AK, Sørensen DB. Strategies for the assessment of competence in laboratory animal science courses. Lab Anim (NY). 2014;43(10):359-63.
8. Laboratory Animal Science Association. Direct Observation of Practical/Procedural skills (DOPS) Available from: https://www.lasa.co.uk/dops/ (accessed 13 April 2022)
9. Education and Training Platform for Laboratory Animal Science (ETPLAS). https://etplas.eu/education/dops-and-sop/ (2018, accessed 13 April 2022)
10. Boateng GO, Neilands TB, Frongillo EA, et al. Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. Frontiers in public health. 2018;6:149.
11. Wang YS, Wang HY, Shee DY. Measuring e-learning systems success in an organizational context: Scale development and validation. Computers in Human Behavior. 2007;23(4):1792-808.

12. Ozkan S, Koseler R. Multi-dimensional students' evaluation of e-learning systems in the higher education context: An empirical investigation. Computers & Education. 2009;53(4):1285-96.
13. Gledhill L, Dale VHM, Powney S, et al. An International Survey of Veterinary Students to Assess Their Use of Online Learning Resources. J Vet Med Educ. 2017;44(4):692-703.
14. Borchers M, Tipold A, Pfarrer C, et al. Acceptance of case-based, interactive e-learning in veterinary medicine on the example of the CASUS system. Tierarztl Prax Ausg K Kleintiere Heimtiere. 2010;38(6):379-88.
15. Bartova E, Halova D. E-Learning for Veterinary students Houska M, Krejci I, Flegl M, editors. Prague 6: Czech University Life Sciences Prague; 2014. 27-31 p.
16. Varnhagen CK, Wright DL. Learning characteristics of veterinary technology students in a distance-education and an on-campus program. J Vet Med Educ. 2008;35(3):449-55.
17. Piccoli G, Ahmad R, Ives B. Web-Based Virtual Learning Environments: A Research Framework and a Preliminary Assessment of Effectiveness in Basic IT Skills Training. MIS Quarterly. 2001;25:401-26.
18. Hong K-S. Relationships between students' and instructional variables with satisfaction and learning from a Web-based course. The Internet and Higher Education. 2002;5(3):267-81.
19. Regmi K, Jones L. A systematic review of the factors – enablers and barriers – affecting e-learning in health sciences education. BMC Medical Education. 2020;20(1):91.
20. Naciri A, Radid M, Kharbach A, et al. E-learning in health professions education during the COVID-19 pandemic: a systematic review. J Educ Eval Health Prof. 2021;18:27.
21. Ibrahim NK, Al Raddadi R, AlDarmasi M, et al. Medical students' acceptance and perceptions of e-learning during the Covid-19 closure time in King Abdulaziz University, Jeddah. Journal of Infection and Public Health. 2021;14(1):17-23.
22. Abdulghani HM, Sattar K, Ahmad T, et al. Association of COVID-19 Pandemic with undergraduate Medical Students' Perceived Stress and Coping. Psychol Res Behav Manag. 2020;13:871-81.
23. Miller GE. The assessment of clinical skills/competence/performance. Academic medicine : journal of the Association of American Medical Colleges. 1990;65(9 Suppl):S63-7.
24. Read EK, Bell C, Rhind S, et al. The use of global rating scales for OSCEs in veterinary medicine. PloS one. 2015;10(3):e0121000.

25. Ma IWY, Zalunardo N, Pachev G, et al. Comparing the use of global rating scale with checklists for the assessment of central venous catheterization skills using simulation. Advances in Health Sciences Education. 2012;17(4):457-70.
26. Cunnington JP, Neville AJ, Norman GR. The risks of thoroughness: Reliability and validity of global ratings and checklists in an OSCE. Advances in health sciences education : theory and practice. 1996;1(3):227-33.
27. Harden RM, Stevenson M, Downie WW, et al. Assessment of clinical competence using objective structured examination. Br Med J. 1975;1(5955):447-51.
28. Wass V, Van der Vleuten C, Shatzer J,et al. Assessment of clinical competence. The Lancet. 2001;357(9260):945-9.
29. Tabish SA. Assessment methods in medical education. Int J Health Sci (Qassim). 2008;2(2):3-7.
30. Preston R, Gratani M, Owens K, et al. Exploring the Impact of Assessment on Medical Students' Learning. Assessment & Evaluation in Higher Education. 2020;45(1):109-24.
31. Bidonde J, Meneses-Echavez JF, Asare B, et al. Developing a tool to assess the skills to perform a health technology assessment. BMC Medical Research Methodology. 2022;22(1):78.
32. Heeneman S, Pool AO, Schuwirth LWT, et al. The impact of programmatic assessment on student learning: theory versus practice. Medical education. 2015;49(5):487-98.
33. Mortaz Hejri S, Jalili M. Standard setting in medical education: fundamental concepts and emerging challenges. Med J Islam Repub Iran. 2014;28:34.
34. De Champlain AF. Standard Setting Methods in Medical Education. Understanding Medical Education2018. p. 347-59.
35. Norcini J, McKinley DW. Concepts in assessment including standard setting. In: Dent JA, Harden RM, Hunt D (editors). A Practical Guide for Medical Teachers. 5 ed. China: Elsevier; 2017. p. 252- 9.
36. Read EK, Farrell R. How Are Clinical Skills Practiced? In: Read ER, Read MR, Baillie S(editors). Veterinary Clinical Skills. 1st ed: Wiley-Blackwell; 2021.
37. Thalheimer W. Measuring Learning Results: Creating fair and valid assessments by considering findings from fundamental learning research. 2007
38. Epstein RM, Hundert EM. Defining and assessing professional competence. Jama. 2002;287(2):226-35.
39. While AE. Competence versus performance: which is more important? Journal of advanced nursing. 1994;20(3):525-31.

8. FINAL REMARKS, IMPLEMENTATION AND FUTURE DIRECTIONS

FINAL REMARKS, IMPLEMENTATION AND FUTURE DIRECTIONS

Education and assessment in Laboratory Animal Science (LAS) is an area that has gained increasing attention in the last decades, with increased understanding of the importance of researcher training to ensure laboratory animals' safety and the quality of research results. So far, little is known about education and assessment strategies in this field and all of these new challenges paved the way for the research presented in this thesis. This thesis findings represents the first systematic research into the topic and the work presented aims to bring insights into different aspects of laboratory animal training and education, with focus on the delivery of knowledge and on assessment methods. These results also support training tutors and institutions in the implementation of programs and the evaluation legally required.

Although limited to the Portuguese context, and to the fact that the assessment studies were performed in one institution, the work herein developed involved more than 400 trainees in laboratory animal science, from around 30 course editions (advanced and introductory), comprising data collected during 7 years, which is illustrative of the studies dimension.

The work reported here can serve as an example for other biomedical institutions that are educating researchers to work properly with animals to develop training programs. We have demonstrated how combining different teaching approaches, namely combining lectures with e-learning, is well accepted and perceived by trainees. E-learning is well accepted by LAS trainees and can be used as a complementary approach to theoretical training.

Regarding assessment in LAS contexts, either in simulation settings or in the workplace, there is a variety of instruments that can be developed to support the exam, such as global rating scales, which proved to "measure" different aspects of trainees´ performance. Practical assessments, such as the OSLASE proposed in this thesis, can be used for formative or summative evaluation and to guide and orient professional development. In their majority, when designed and implemented adequately, assessments are known to impact the development of trainees' competences positively, highlighting the importance of proper development of practical skills. Institutions should design their programs considering the inclusion of evaluation/competence verification at different timings to assess representative skills, according future researcher functions and responsibilities.

Above all, to ensure high standards of animal welfare and safety, scientific research in LAS, and fulfil what is required by Directive 2010/63/EU regarding education and competence, it is essential that European biomedical institutions develop and implement training programs which include practical assessments. Designing and implementing assessments of practical competence is not simple and it is a continuous process of development and exploration of the instrument limitations and the exam strategy in practice. Again, the work presented in this thesis can serve as an example for other institutions in this regard.

Finally, it is pressing to establish a wider discussion in the field about what should be assessed, when, where and by whom. In particular, it is urgent to define what is meant by practical competence and proficiency in LAS. The creation of a common repository where tested and verified assessment instruments might be shared with all LAS community could be the first step to disseminate and implement standards in the field.

APPENDICES

APPENDIX I
Questionnaires to assess students' acceptance of e-learning approaches in laboratory animal science training

Questionnaires administered to conduct the study reported in chapter 4.

Questionnaire of E-learning Acceptance

1) Age: ____

2) Gender: ☐ Male ☐ Female

3) Highest level of education already completed:

☐Bachelor ☐ Master ☐ PhD. ☐ Other:_______

4) Area of your first completed university degree:

☐ Biology ☐ Pharmacy ☐ Biochemistry ☐ Veterinary ☐Veterinary Nursing ☐ Medicine

☐ Other:_______

5) Position/ Current occupation:

☐Technician ☐ Undergraduate training ☐ PhD student ☐ Post-doc ☐PI ☐ Other:_______

6) Is this your first time using e-learning platform?☐ no ☐ yes

7) How many working hours per week were you expecting to invest in this course?

☐ < 7 hours/week ☐ 7- 35 hours/ week ☐ 35- 70 hours/ week ☐ > 70 hours/week

8) When starting this course, which were your expectations about difficulty in using e-learning?

☐Low ☐High ☐Very high ☐I did not know what to expect

9) When starting this course, which were your expectations about efficacy in using e-learning?

☐Low ☐High ☐Very high ☐I did not know what to expect

10) When starting this course, did you have fears of laboratory animals? ☐no☐ yes

11) When starting this course, how much previous experience did you have of working with laboratory animals? ☐none ☐ less than 1 year ☐ 1 year or more

12) When did you start using the e-learning platform?

☐before the start of the course ☐ at the start of the course

The following table presents you with a number of statements. Please indicate how well you agree with each of them by circling the number that best fits with your opinion.

1 Strongly disagree	2 disagree	3 neutral	4 agree	5 Strongly agree

	1	2	3	4	5
1. The e-learning platform seems to be updated regularly	1	2	3	4	5
2. The e-learning platform works well .	1	2	3	4	5
3. The e-learning platform is user-friendly	1	2	3	4	5
4. The e-learning platform has attractive features that appeal to the users	1	2	3	4	5
5. The e-learning platform allows a more interactive learning	1	2	3	4	5
6. The different e-learning activities are designed to get the best out of students	1	2	3	4	5

7. Completing the e-lessons made me well prepared for the classroom lectures	1	2	3	4	5
8. Online quizzes helped me to learn effectively	1	2	3	4	5
9. Videos/animations were used appropriately in the e-learning activities	1	2	3	4	5
10. Images/graphics were used appropriately in the e-learning activities	1	2	3	4	5
11. The e-learning activities provide accurate information	1	2	3	4	5
12. The e-lesson quizzes are easy to understand	1	2	3	4	5
13. When I needed help to use the e-learning platform I knew where to find it	1	2	3	4	5
14. I was given access to clear instructions about how to use the e-learning before starting	1	2	3	4	5
15. The e-lessons are well structured and are easy to follow	1	2	3	4	5
16. The contents are organized in an intuitive and easy way	1	2	3	4	5
17. The e-lessons are organized in a progressive way, starting with basics and progressing to more complex issues	1	2	3	4	5
18. My experience with e-learning has been enjoyable.	1	2	3	4	5
19. I would recommend this e-learning to other students	1	2	3	4	5
20. Combining two ways of presenting information (lectures and e-learning) was helpful for my learning process	1	2	3	4	5
21. The e-lesson quizzes helped me to integrate information and consolidate knowledge	1	2	3	4	5
22. I found the dynamic of the e-lessons appropriate to keep me motivated	1	2	3	4	5
23. In this course it is convenient that all information (e g news, handouts, complementary literature) is available through the e-learning	1	2	3	4	5
24. Presenting part of the content through e-learning is a good solution to reduce the intensity during the course duration	1	2	3	4	5
25. The e-learning system enables me to control my learning progress, and invest more time in topics I find more difficult	1	2	3	4	5
26. The e-learning system enables me to invest more time in topics that interest me more	1	2	3	4	5
27. By presenting part of the content as e-learning the schedule becomes more flexible for the student	1	2	3	4	5
28. The videos shown in e-learning helped me to prepare for practical classes of animal handling	1	2	3	4	5
29. The videos shown in e-learning helped me to prepare for practical classes of animal procedures	1	2	3	4	5
30. The videos shown in e-learning helped me to prepare for practical classes of animal necropsy	1	2	3	4	5
31. The videos shown in e-learning lectures helped me to deal with my possible animal fears	1	2	3	4	5
32. The several videos in the e-learning platform helped me to get a better understanding of how animals behave and react that is useful for me when handling animals	1	2	3	4	5

SHADOW ITEMS DID NOT INTEGRATE THE VERSION ADMINISTERED TO STUDENTS IN THE INTRODUCTORY COURSE.

APPENDIX II
Global rating scales to assess the competence of researchers in handling and performing substance administrations in laboratory rodents

Scales developed in the context of the work reported in chapter 5.
Published at the Open Repository of the University of Porto.

Global rating scales to assess the competence of researchers in handling and performing substance administrations in laboratory rodents

Andreia Costa[1,2,3], Sofia Lamas[1,2], Manuel João Costa[4], I. Anna S. Olsson[1,2]

1 i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal
2 IBMC - Institute for Molecular and Cell Biology, University of Porto, Portugal
3 ICBAS - Abel Salazar Institute of Biomedical Sciences, University of Porto, Portugal
4 ICVS - Life and Health Sciences Research Institute, School of Medicine, University of Minho, Portugal

Index

Instructions to the assessor

1. Ensure that the candidate's identification (name and course/assessment reference) and the date of task performance are correct.
2. Provide clear indications to the candidate regarding the task required at each assessment (see the instructions to the candidate indicated in the different global rating scales).
3. Ensure that the candidate is wearing appropriate protective equipment before starting the demonstration.
4. Observe the candidate without interfering or communicating, except if the candidate expresses attitudes or behaviors that might compromise animal well-being, in which case the demonstration should be immediately halted, and the candidate's performance should be considered "not approved".
5. Use the scale to assess the candidate's performance by selecting for each descriptor the categories representing the candidate's actions.
6. Feedback may be given to candidate with focus on both the positive and negative aspects of the demonstration, to promote skills development.
7. Please note that information that is considered relevant, such as any special condition of the station that constrains the candidate's performance, the animal strains used, disturbances occurring during the assessment, etc., should be registered under the "Observations" column.

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

Suggested approval criteria for summative examinations

- We recommend adjusting the approval criteria to the context of the assessment, considering institutional standards and policies, the objectives of the assessment, and the intended proficiency of the candidates. Such criteria should be defined prior to examinations and be transparent to candidates.
- The pass criteria recommended for candidates who concluded formal training in laboratory animal science is a final score of [pass] for “global impression”, with at least one [clear pass] record for all the remaining items.
- We propose that these pass criteria correspond to level 2/3 of supervision, as defined in the EU working document[1]. This score indicates that, although the candidate developed the skill, training with supervision is required to achieve mastery (level 0 of supervision, defined in the EU working document [1])

---

[1] "The level of supervision can be divided in five stages:
4 – Supervisor present when the procedure takes place providing direct supervision and advice
3 – Supervisor aware when procedures are taking place and available for rapid intervention if required (i.e. in the vicinity of the procedure)
2 – Supervisor aware when procedures are taking place and available to attend to provide advice if required (i.e. in the vicinity of the establishment)
1 – Supervisor aware when procedures are taking place and available for discussion to provide advice if necessary (e.g. by telephone)
0 – No supervision required"

European Commission guidance document. National Competent Authorities for the implementation of Directive 2010/63/EU on the protection of animals used for scientific purposes. A working document on the development of a common education and training framework to fulfil the requirements under the Directive. Brussels, http://ec.europa.eu/environment/chemicals/lab_animals/pdf/guidance/education_training/en.pdf (2014, accessed 9 April 2021). page 14

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

Mouse restraint

Candidate Name:

ID/ Course reference:

Assessor Name:

Assessment Station ID:

Instructions to the candidate:

1. Please remove one animal from the cage.
2. Perform the complete restraint.
3. Return the animal safely to the cage.

MOUSE Rt		FAIL	PASS	CLEAR PASS
Removal from the cage	Tail grasp attempts	Failed or >8	3–7	<3 or tunnel
	Suspension	>10 s	5-10 s	<5 or tunnel
	Tail hold	Failed	tip or middle	Base
Restraint & scruffing	Neck scruff attempts	Failed or >10	3-10	<3
	Scruffing	No hold or too loose/too tight	Loose/tight	Adequate
	Tail holding/restraint	Failed or posterior part of the animal released	Tail hold via other way than with the ring finger and/or without the thumb supporting the body	Adequate animal body support using the thumb
After restraint	Procedure repetition	>3	1–3	No repetition
GLOBAL IMPRESSION		☹	😐	☺

OBSERVATIONS:

Final score: ☐ fail ☐ pass ☐ clear pass

Assessor signature:

Date:

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

Rat restraint

Candidate Name: Assessor Name:

ID/ Course reference: Assessment Station ID:

Instructions to the candidate:
1. Please remove one animal from the cage.
2. Perform the complete restraint.
3. Return the animal safely to the cage.

RAT Rt 		FAIL	PASS	CLEAR PASS
Removal from cage	Removal from the cage and animal suspension	Tail (if not necessary)	Quickly by the tail (>2 s)	Thorax or tunnel
	Holding the animal (in the operator's arms)	Animal free in the arm/not holding	Not always holding	Always holding (by the tail or body)
Restraint & scruffing	Cranial/anterior restraint (head and jaw)	Animal can move head/not holding or suffocating animal	Holding loosely or tight/hiding thorax	Adequate
	Caudal/posterior restraint (legs and paws)	Animal can move/not holding	Holding loosely or tight/hiding abdomen	Adequate
After restraint	Return to the cage	Tail or dropped	Quickly by the tail (>2 sec)	Thorax or tunnel, or the animal walked down into the cage
	Procedure repetition	>3	1-3	No repetition
GLOBAL IMPRESSION				

OBSERVATIONS:

Final score: □ fail □ pass □ clear pass

Assessor signature: Date:

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

Mouse intraperitoneal administration

Candidate Name: Assessor Name:

ID/ Course reference: Assessment Station ID:

Instructions to the candidate:

1. Please remove one animal from the cage,
2. Perform an intraperitoneal administration using the material in the station.
3. Return the animal safely to the cage.

MOUSE IP		FAIL	PASS	CLEAR PASS
Before admin.	Restraint	Failed/no hold	Loose/tight	Adequate
	Animal position	Head up	Parallel	Head down
Administration	Body part (quadrant)	Wrong, outside margins	Quadrant margins	Correct (inside quadrant)
	Depth of puncture angle	Depth: superficial or deep Angle:	Depth: appropriate Angle:	Depth: appropriate Angle:
	Reflux/syringe holding	No reflux	Yes, reflux/trembling	Yes, reflux/holding firmly
	Injection speed	Failed too slow /too fast	Slow/fast	Acceptable
After admin.	Time after restraint	> 60 s	60-30 s	< 30 s
	Procedure repetition	> 3	1-3	No repetition
GLOBAL IMPRESSION				

OBSERVATIONS:

Final score: ☐ fail ☐ pass ☐ clear pass

Assessor signature: Date:

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

Rat subcutaneous administration

Candidate Name:

Assessor Name:

ID/ Course reference:

Assessment Station ID:

Instructions to the candidate:

1. Please remove one animal from the cage,
2. Perform a subcutaneous administration using the material in the station.
3. Return the animal safely to the cage.

RAT SC		FAIL	PASS	CLEAR PASS
Before admin.	Restraint: animal body orientation & puncture site	Body: failed or perpendicular orientation Puncture site: not visible	Body: no parallel orientation Puncture site: visible	Body: parallel orientation Puncture site: visible
Administration	Body part (quadrant)	Wrong	Acceptable but not into scapula	Scapula
	Depth of puncture angle	Depth: superficial or deep Angle:	Angle: Depth: appropriate	Depth: appropriate Angle:
	Reflux/syringe holding	No reflux	Yes, reflux/trembling	Yes, reflux/holding firmly
	Injection speed	Failed too slow /too fast	Slow/fast	Appropriate
After admin.	Time after restraint	> 60 s	60–30 s	< 30 s
	Procedure repetition	> 3	1–3	No repetition
GLOBAL IMPRESSION				

OBSERVATIONS:

Final score: ☐ fail ☐ pass ☐ clear pass

Assessor signature:

Date:

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

Stations material

Mouse restraint	cage with lid tunnel
Rat restraint	tunnel
Mouse intraperitoneal administration	Sterile syringe (1 ml) Sterile needle (25 – 27 G; 5/8" length or less) Saline for injection Sharps container
Rat subcutaneous administration	Sterile syringe (1 - 5 ml) Sterile needle (23 - 25G; 5/8" length or less) Saline for injected Sharps container

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

APPENDIX III
The development of a high stakes OSCE to enhance the assessment of practical competence in Laboratory Animal Science Courses

Published abstract: Poster communication (P.12) at Annual European Conferenceon Assessment in Medical Education 'Technology Enabled Assessment' (EACCME) Costa et al. Braga, Portugal 2018.

The development of a high stakes OSCE to enhance the assessment of practical competence in Laboratory Animal Science Courses

Main Author: Andreia Costa
i3S/University of Porto (Portugal)
afcosta@ibmc.up.pt

Researchers carrying out procedures with living animals are legally required to undergo training to develop competence to handle animals appropriately. Common standards for training and competence assessment have recently been introduced to harmonize training internationally. Practical competence is necessary before researchers do unsupervised work in laboratory animal facilities and ensuring that new animal users are properly trained is both a moral requisite and a legal requirement under the European Directive 2010/63/EU. However, a valid and reliable assessment of practical competence at the conclusion of animal training has not been developed. Here, we present the development of an ongoing project to create an Objective Structured Clinical Exam for practical skills in laboratory animal science (LAS).

The purpose of this study was to develop an objective assessment instrument to attest practical student ability to perform handling and procedures with rodents, namely, mouse restraint, rat restraint, mouse intraperitoneal injection (ip) and rat subcutaneous injection. Particularly, here, we will present the process of developing the instrument for mouse ip evaluation. The ip technique was chosen for being an essential and worldwide used procedure in biomedical research with rodents. It is also one of the procedures taught in all the LAS courses that run regularly at our institution. Despite its importance, there is yet no marking rubric available to use in LAS education to assess practical competence.

The process of development and definition of the marking rubric was initially based on revision of the standard reference material and experts' contribution – trainers in practical classes, technicians and the responsible laboratory animal veterinarian. Within this process, measurable and crucial steps for the success of the ip procedure were identified and a consensus for trainee approval criteria was established. Subsequently, the instrument was tested in a pilot evaluation context, to check the items adequacy as well as the operationally of its application in the OSCE setting.

The third step in this development process was the incorporation of a number of changes based on specialists' revision; two items were eliminated and two of the four categories condensed (fail and poor). A final score sheet was established comprising 9 items (corresponding to 8 tasks plus 1 item of global impression) with a 3-point answering scale (from fail to clear pass).

Globally, the preliminary results presented a high level of inter-observer reliability (1 _ intraclass correlation coefficient scores _ 0,6) and adequacy to a OSCE in LAS context. Also, the main limitations and issues faced in the establishment of the practical competencies' evaluation within this context will be discussed.

APPENDIX IV
Laboratory Animal Science online training through the eyes of the learner: perceptions of gains

Published abstract: Oral communication (OA4S2) at 14th FELASA Congress 2019
Costa et al. Lab Anim 2019 53(1S) pp.31.

OA4S2

Laboratory Animal Science online training through the eyes of the learner: perceptions of gains

Costa Andreia[1,2], Costa A.[3] and Olsson I.A.[1]
[1]*Laboratory Animal Science group, i3S, Instituto de Investigação e Inovação em Saúde – University of Porto / IBMC, Institute for Molecular and Cell Biology, University of Porto, Porto, Portugal*
[2]*ICBAS, Abel Salazar Institute of Biomedical Sciences, University of Porto, Porto, Portugal*
[3]*FPCEUP, Faculty of Psychology and Education Sciences of the University of Porto, Porto, Portugal*

Abstract
The European directive 2010/63/EU issued clear recommendations concerning the training and competence of researchers using animals in experimental procedures. Different courses and training programs have emerged across Europe, and the Laboratory Animal Science community is facing the challenge of offering qualified and flexible training. From similar educational areas, E-learning appears to be a promising solution tool, fulfilling the flexible-time training and quality criteria.

However, very few studies have addressed the participants' perception of e-learning as a training tool. For course organizers the e-learning approach is the key to the problem of demanding programs and short training periods. However, participants' perceptions, that is known to be a relevant factor for learning have not been investigated.

Within an exploratory study framework, the e-learning approach perception of 229 researchers participating in LAS training programs was assessed. Participants answered the Questionnaire of E-learning Acceptance (QELA), a Likert-type scale comprising 5 subscales that explores how participant perceive the organization of the e-learning, its relevance for the time management and for preparing for practical training. Results also explored the influence of student's sociodemographic and professional profile in the e-learning acceptance.

In general, e-learning was students' well accepted by the majority of the participants as an alternative to classroom lectures. We will further discuss how we have used the results to improve our use of e-learning.

APPENDIX V
Are you skilled enough to work with animals? Measuring competencies using OSCE stations

Published abstract: Workshop organization (OA12W1) at 14th FELASA Congress 2019
Costa et al. Lab Anim 2019 53(1S) pp. 41

OA12W1

Are you skilled enough to work with animals? Measuring competencies using OSCE stations

Costa A and Olsson Anna
i3S – Institute for Research and Innovation in Health, University of Porto, Porto, Portugal

Abstract
The terms competence/competent are used in Directive 2010/63/EU as a requirement for persons carrying out procedures with animals. More recently, several initiatives have started to address the question of competence is to be assessed in order to guarantee that future animal users are skilled enough to be authorized to work without supervision.

Following the Objective Structured Clinical Exam (OSCE) approach, we have developed an instrument to assess students' ability to perform handling and key procedures with rodents (restraint and two types of injections).

The process of development and definition of the marking rubric was initially based on revision of the standard reference material and experts' contribution – trainers in practical classes, technicians and the responsible laboratory animal veterinarian. Within this process, measurable and crucial steps for the success of the procedures were identified and a consensus for approval criteria was established.

Subsequently, the instrument was tested in a pilot evaluation context, to check that the itmes were adequate and operational in a practical setting.

The third step in this development process was the incorporation of a number of changes based on specialists' revision; items were eliminated and categories condensed and a final score sheet for each procedure was established with a 3-point answering scale (from *fail* to *clear pass*).

In this workshop, participants will be given an introduction to the concept of OSCE and how to develop them, and thereafter get the opportunity to practice scoring different student performances from video recordings.

APPENDIX VI
An Objective Structured Laboratory Animal Science Examination (OSLASE) to ensure researchers' professional competence in laboratory animal science

Published in Laboratory Animals (2022), 0 (0)
doi:10.1177/00236772221135671

Laboratory Animals Limited

Special Issue: Education

Laboratory Animals
0(0) 1–11

DOI: 10.1177/00236772221135671
journals.sagepub.com/home/lan
SAGE

An Objective Structured Laboratory Animal Science Examination (OSLASE) to ensure researchers' professional competence in laboratory animal science

Andreia Costa[1,2,3], Sofia Lamas[1,2], Maria Rui Correia[1,2], Maria S Gomes[1,2,3], Manuel J Costa[4] and I Anna S Olsson[1,2]

Abstract

The evaluation of the competence of personnel working with laboratory animals is currently a challenge. Directive 2010/63/EU establishes that staff must have demonstrated competence before they perform unsupervised work with living animals. Nevertheless, there is a lack of research into education and training in laboratory animal science, and the establishment of assessment strategies to confirm researchers' competence remains largely unaddressed.

In this study, we analysed the implementation of a practical assessment strategy over three consecutive years (2018–2021) using the Objective Structured Laboratory Animal Science Exam (OSLASE) developed previously by us to assess professional competence. The interrater reliability (IRR) was determined based on the assessors' rating of candidates' performance at different OSLASE stations using weighted kappa (Kw) and percentage of agreement. Focus group interviews were conducted to access trainees' acceptability regarding the OSLASE.

There was a moderate-to-good Kw for the majority of the scales' items ($0.79 \pm 0.20 \leq \text{Kw} \geq 0.45 \pm 0.13$). The percentages of agreement were also acceptable ($\geq$75%) for all scale items but one. Trainees reported that the OSLASE had a positive impact on their engagement during practical training, and that it clarified the standards established for their performance and the skills that required improvement. These preliminary results illustrate how assessment strategies, such as the OSLASE, can be implemented in a manner that is useful for both assessors and trainees.

Examen structuré objectif de science animale de laboratoire (OSASSE) pour assurer la compétence professionnelle des chercheurs en SAL

Keywords

Competence and proficiency, performance assessment, Objective Structured Laboratory Animal Science Exam (OSLASE), LAS assessment impact, education and training in laboratory animal science

Date received: 6 May 2022; accepted: 12 October 2022

Driven by an increased awareness of the importance of ensuring researchers' professional competence in animal experimentation to safeguard animal welfare, the interest in the design and implementation of innovative assessment methods has been increasing notoriously. Such an increase is motivated not only by the imposition of the revised European legislation,[1,2] but also by the need to guarantee the application of the 3Rs in practice.[3]

Directive 2010/63/EU establishes that staff working with living animals must be 'adequately educated and trained' and 'supervised in the performance of their

[1]i3S, Instituto de Investigação e Inovação em Saúde, University of Porto, Portugal
[2]IBMC, Institute for Molecular and Cell Biology, University of Porto, Portugal
[3]ICBAS, Abel Salazar Institute of Biomedical Sciences, University of Porto, Portugal
[4]ICVS, Life and Health Sciences Research Institute, School of Medicine, University of Minho, Portugal

Corresponding author:
Andreia Costa, CostaInstituto de Investigação e Inovação em Saúde, University of Porto Rua Alfredo Allen 208 4200-135 Porto, Portugal.
Email: afcosta@ibmc.up.pt

tasks until they have demonstrated the requisite competence'.[2] The associated working document for a common education and training framework further clarifies that the assessment of practical competence should rely on observations of trainees while they perform the procedures.[1] Such regulations call out for greater awareness by education and training supervisors regarding competence and assessment constructs in the laboratory animal science (LAS) context.[4] Ensuring professional competence in LAS through observational assessments is critical at different stages of learning; for example, as a final examination at the conclusion of a LAS training period, before the beginning of autonomous work, or as a formative work-integrated assessment as part of continuing professional development (CPD).

The field of LAS can draw on the extensive knowledge regarding the assessment of professional competence developed in the context of medical and health professional education. In these fields, driven by specific professional regulations and the agenda of patient safety, concepts such as competence, performance, and the validity and reliability of the assessment methods, have been largely explored.[5–7] Assessments are crucial elements in the educational process and play a relevant role in all phases of professional development.[8] Moreover, assessment is a key element in the process of learning, having a positive impact and affecting learners in regard to how they plan and carry out their studies.[6–9]

Regardless of the timing of their application, the format of observational assessments is a crucial element for ensuring that they are capable of measuring what is intended. In LAS, there is a need for the observational measurement of professional performance other than cognitive knowledge, which is suitably assessed using traditional formats, such as written or oral tests. In medical and health science education, there is consolidated experience in alternative formats of observational exams, such as the Objective Structured Clinical Examination (OSCE), the mini-clinical evaluation, simulations and direct observation in clinical workplaces.[6,8,10] In this sense, the format and design of the assessments must be planned to allow the capture of psychomotor dimensions (technical ability) and, if possible, emotional/attitudinal dimensions (such as professionalism or empathy with animals).[11,12]

Although LAS training programmes are increasingly focusing on professional competence in practice, assessments focused on abilities in this domain are still lacking. LASA was the first organization to publish instruments to assess practical competence in LAS, namely the Direct Observation of Practical/Procedural Skills (DOPS).[13] More recently (starting in late 2021), the Education and Training Platform for Laboratory Animal Science (ETPLAS) website has been disseminating DOPS sheets developed by the working groups, to be used as examples in assessments.[14]

As part of an ongoing project implemented at our institution aimed at developing an Objective Structured Laboratory Animal Science Exam (OSLASE) of researchers' competence, we have previously developed instruments to capture the performance of researchers while handling and performing substance administrations in laboratory rodents.[11,15] In this study, we continued the development process of this tool by evaluating the interrater reliability of the assessors and the acceptability of the OSLASE for those who have taken the examination. These are two important elements to understand better the relevance and the assessment impact . Interrater reliability is critical for the reliability of any assessment. As the OSLASE is quite different from the existing examinations, understanding how participants experience the examination should reveal its impact on their motivation and engagement, exam preparation, and development of competence.

Material and methods

Study design and setting

This study included participants who completed a LAS training course (see Table 1 for details of the two course formats). The OSLASE is an observational examination designed to assess LAS skills performance and competence inspired by the OSCE in health professional education.[11,16] The exam consisted of four manned stations which assessed the performance of the participants during essential and routine procedures using laboratory rodents; that is, mouse restraint, rat restraint, mouse intraperitoneal administration and rat subcutaneous administration. All stations were timed and involved the handling and restraint of a live rodent (mouse restraint and rat restraint), as well as the execution of procedures using mannequins (mouse intraperitoneal administration and rat subcutaneous administration). Table 1 depicts the stations that integrated the OSLASE in the two course formats. In both contexts (introductory and advanced courses), performance was assessed by trained examiners supported by a global rating scale (GRS) specifically developed for these procedures.[15]

Data from OSLASE assessments were collected over three years, from 2018 to 2021.

At our institution, a pass mark in the OSLASE (practical assessment) is necessary to successfully conclude the LAS course. Those who fail need to retake the practical training. In the introductory course, trainees take the OSLASE and a written assessment at the end of training. A choice of dates for the OSLASE is

Table 1. Introductory and advanced training courses: description and main assessment aspects.

	Introductory course	Advanced course
	Covers Directive Function A and D Spaced in time (approximately two months) ≥40 h	Covers Directive Function A, B and D Condensed in time (10 full days) ≥80 h
Assessment timing	Final (at the end of practical training)	During training process (between practical course modules, at the end of handling module)
OSLASE stations	mouse restraint rat restraint mouse intraperitoneal administration rat subcutaneous administration	mouse restraint rat restraint
Data collected	interrater agreement trainees exam acceptability	trainees exam acceptability

OSLASE: Objective Structured Laboratory Animal Science Exam.

offered between two options spaced approximately by one month for the introductory training. In the advanced course, the OSLASE is integrated in the rodent handling practical module, and participants can choose to undertake it at any time during the two sessions of this module. At both training levels, the trainees are assessed individually in a standardized scenario by a trained assessor. The main aspects of each of the training courses are summarized in Table 1.

Data collection from human participants (instrument application in the context of the courses and group interviews) was performed under approval from the Joint Ethics Committee of CHUP/ICBAS with reference numbers 2019/CE/P014(292/CETI/ICBAS) and 2021/CE/P013(P353/CETI/ICBAS). Informed consent was obtained from the candidates prior to data collection. The use of animals for teaching was covered by approval 2017-03 by the i3S Animal Welfare and Ethics Review Body.

OSLASE set-up

In each station, the assessors used a specific global rating composed of 7 to 9 items with descriptors. The final item of each scale was the 'global impression' item. The items of all scales presented three performance-rating categories: fail, pass, and clear pass.

In introductory LAS courses, the candidates were assessed individually at the end of the training by the first and second authors of this study. The two assessors evaluated each candidate ($n = 92$) independently, at the same time, via direct observation of their performance. Not all candidates included in the sample completed the four stations. Some candidates failed scale items, which led to performance failing/end of the exam (e.g. failing to pick up the animal from the cage). Furthermore, a few candidates received training for only one of the species; therefore, they were assessed only for that species (e.g. when training for work using mice, they performed mouse restraint and mouse intraperitoneal administration).

Evaluation of interrater agreement

The evaluation of the OSLASE for the purpose of this paper included assessing the agreement between the ratings of the two assessors, that is, the interrater reliability (IRR), in two ways: the percentage of absolute agreement (% of agreement) and the weighted kappa (Kw) statistic. These are recommended methods for determining assessor agreement because of the recognition of the importance of decisions in healthcare and clinical research supported by IRR results.[17] The percentage of absolute agreement between the two assessors was the number of concordant scores divided by the total number of scores. A range from 75% to 90% of absolute agreement is considered to be acceptable.[17,18] The Kw, with linear weighting, was calculated for all scale items of the four global rating scales and for the global impression item (the last item of each scale). Kw allows the weighting of disagreements differently, which is especially useful in studies applying scales with ordered rating categories, so that it confers greater emphasis to large differences between ratings than it does to small differences.[19–21] Kappa values representing the agreement between raters were interpreted according to the following scheme: <0.20, poor; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, good; and 0.81–1.00, very good.[22] The analyses were performed using the SPSS 21 statistical software package (SPSS Inc., Chicago, USA).

Evaluation of acceptability

To capture a variety of perceptions in both course levels concerning the acceptability of the OSLASE, thirteen LAS course trainees (seven participants from the advanced course, two different editions and the remaining from the introductory course, three different courses) participated in the focus group interviews. Eligible trainees had successfully completed the course at either level (introductory or advanced) in the preceding twenty four months and had received a certificate. Participants were invited by email. All respondents who accepted the invitation were integrated in the interviews. They were previously informed about the study aim and conditions and gave their consent to participate. The decision to conduct focus groups envisioned a deeper delving into trainees' views. The interview guide was developed by the study authors and was adapted to the different courses (see Table 1). Because of COVID-19 restrictions, the interviews were carried out online (via Zoom).

A total of four focus group interviews were conducted over a seven-month period. Trainees were allocated to focus groups according to course type (two focus groups per type) and availability. The interviews lasted approximately 90 min, with an initial introduction followed by a moderator (last author) - facilitated discussion; an assistant moderator (third author) took notes during the interviews. Neither the moderator nor the assistant moderator were related to the OSLASE administration. The participants were first provided an introduction about the purpose of the study, followed by script questions. The participants discussed their own experiences freely, in combination with prompts when necessary. With the participants' permission, all sessions were audio and video recorded and the content was transcribed verbatim. Participant identification was pseudonymized. The verbatim transcripts were coded using the NVivo software (QSR International 2022) version 12 for qualitative analysis. The first and last authors decided on a first set of codes and coded each transcription separately, before meeting to confirm the coding and revise the citations. At the end of the procedure, five categories were identified (Choice of timing to perform the exam; Effect on trainee practical preparation before performing the OSLASE; Experience and perception; OSLASE impact on trainee learning process; and OSLASE relevance and importance).

Results

Determining the agreement between raters

The mouse restraint scale items related to 'tail grasp attempts', 'tail hold', and 'suspension duration' showed high percentages of agreement between raters ($\geq$85%), whereas the Kw revealed poor agreement (below 0.40). The remaining items of this scale showed moderate-to-good Kw values ($0.45 \pm 0.13 \leq \mathrm{Kw} \geq 0.61 \pm 0.08$). The lowest percentage of agreement for all items of this scale was obtained for the 'global impression' item (71%), although it presented a moderate Kw value (0.51 ± 0.07) (Table 2, part A).

Regarding the rat restraint scale, the Kw values for interrater agreement for the items 'hold the animal' (0.32 ± 0.24) and 'return to the cage' (0.27 ± 0.24) were low, despite the fact that the same items exhibited higher percentages of agreement (>95%). For the remaining scale items, including the 'global impression' item, there was moderate interrater agreement ($0.48 \pm 0.11 \leq \mathrm{Kw} \geq 0.56 \pm 0.12$), with percentages ranging from 76% to 88%. The first scale item, 'remove the animal from cage', showed a good coefficient ($\mathrm{Kw} = 0.75 \pm 0.12$) and percentage of agreement (97%) (Table 2, part B).

Regarding the mouse intraperitoneal administration scale, it was not statistically possible to calculate kappa values for the restraint and injection speed items. The percentages of agreement were acceptable ($\geq$97%). One item on this scale, 'body part/quadrant', exhibited a low interrater agreement ($\mathrm{Kw} = 0.33 \pm 0.15$) and a high percentage of agreement (90%). The remaining scale items showed moderate ('depth of puncture/angle' and 'global impression'), good ('animal position' and 'reflux/syringe holding') and excellent ('animal position' and 'time after restraint') kappa values, with percentages of agreement ranging between 75% and 100% (Table 2, part C).

For some items on the rat subcutaneous administration scale ('body part/quadrant', 'injection speed', 'time after restraint,' and 'procedure repetitions'), similarly to intraperitoneal administration, it was not possible to calculate kappa because of the statistical assumptions underlying the Kw calculation. Nevertheless, those same items showed acceptable percentages of agreement (96%–100%). The kappa of the remaining items was moderate ('depth of puncture/angle' and 'global impression') and good ('restraint/animal position' and 'reflux/syringe holding'). The percentages of agreement for these scale items ranged between 83% and 95% (Table 2, part D).

The 'global impression' item of each scale exhibited moderate values of kappa and an acceptable level of percentage of agreement: global impression for the mouse restraint scale, $\mathrm{Kw} = 0.51 \pm 0.07$ and 71%; global impression for the rat restraint scale, $\mathrm{Kw} = 0.49 \pm 0.10$ and 76%; global impression for the mouse intraperitoneal scale, $\mathrm{Kw} = 0.57 \pm 0.09$ and 75%; and global impression for the rat subcutaneous administration scale, $\mathrm{Kw} = 0.59 \pm 0.11$ and 83% (Table 2).

Table 2. Calculation of percentage of agreement (% agreement) and weighted kappa (Kw) between raters for all items on the four different global rating scales, that is, (A) Mouse restraint, (B) Rat restraint, (C) Mouse intraperitoneal administration, and (D) Rat subcutaneous administration, for the total number of performances (n performances).

Scale item		*n* performances	% agreement	Kw	SD
A | *Mouse restraint*					
Pick from the cage	Tail grasp attempts	92	96	−0.02	0.01
	Tail hold	92	85	0.18	0.12
	Suspension duration	89	96	0.31	0.25
Restraint and scruffing	Neck scruff attempts	89	78	0.54	0.09
	Scruffing	92	76	0.51	0.09
	Tail holding/restraint	92	85	0.45	0.13
After restraint	Procedure repetition	92	75	0.61	0.08
Global impression		92	71	0.51	0.07
B | *Rat restraint*					
Pick from the cage	Remove from the cage	88	97	0.75	0.12
	Hold the animal	88	95	0.32	0.24
Restraint and scruffing	Cranial/anterior restraint	88	82	0.48	0.11
	Caudal/posterior restraint	88	88	0.56	0.12
After restraint	Return to the cage	88	97	0.27	0.24
	Procedure repetition	88	78	0.55	0.10
Global impression		87	76	0.49	0.10
C | *Mouse intraperitoneal administration*					
Before administration	Restraint	88	97	[a]	[a]
	Animal position	88	92	0.90	0.04
Administration	Body part/quadrant	88	90	0.33	0.15
	Depth of puncture/angle	87	75	0.54	0.08
	Reflux/syringe holding	88	80	0.74	0.09
	Injection speed	88	99	[a]	[a]
After administration	Time after restraint	88	100	1	0
	Procedure repetition	88	99	0.79	0.20
Global impression		88	75	0.57	0.09
D | *Rat subcutaneous administration*					
Before administration	Restraint/animal position	84	95	0.82	0.9
	Body part/quadrant	84	98	[a]	[a]
Administration	Depth of puncture/angle	84	89	0.59	0.13
	Reflux/syringe holding	84	89	0.71	0.10
	Injection speed	82	100	[a]	[a]
	Time after restraint	83	96	[a]	[a]
After administration	Procedures repetitions	84	100	b	
Global impression		84	83	0.59	0.11

[a]One of the variables was constant; therefore, Kw could not be computed in this instance.
[b]Each variable in each pair had less than two valid categories; therefore, Kw could not be computed.

OSLASE acceptability among the trainees

We considered that we reached saturation (i.e. that new interviews no longer added perspectives that had not emerged in previous interviews) in the four focus groups, in which both course types (introductory and advanced) were represented. From the focus group interviews performed among former trainees, five main themes emerged related to the practical exam. Verbatim quotes for each category are given as examples in Table 3. Respondent pseudonym and course format are indicated at the end of each quote.

(1) Choice of timing

When the advanced course participants were asked about the factors that determined their choice of timing to perform the OSLASE, the key factors were feeling confident and making the best use of the available time to practice before the exam. Without exception, interviewees from the introductory course reported having chosen to take the exam after they had concluded the training period with their tutor, rather than directly after the end of the practical classes. These students, whose practical training had been impacted by restrictions during the COVID-19 pandemic, also highlighted the balance between avoiding a too-long gap but still having sufficient time to practice.

(2) Effect on trainee practical preparation before performing the OSLASE

Several of the interviewees from the advanced course stated that the existence of a formal practical exam enhanced their engagement in the practical class of animal handling. However, some participants argued that they were not approaching the learning any differently as a consequence of knowing they would have to pass an exam. Among the interviewees from the introductory course, there was also a widespread view that knowing there would be a formal practical exam had little influence on how they prepared and practiced. Students also mentioned that, because they were preparing for real life work with live animals, their preparation was not dependent on the administration of an exam.

(3) Experience and perception

Some respondents reported that they unavoidably felt anxious at having to perform a practical demonstration, whereas others highlighted the fact that they did not feel much pressure, as a consequence of how they had prepared and how the exam was carried out, with few trainees per instructor.

(4) OSLASE impact on the trainee learning process

Among the interviewees from the advanced course, in whom the OSLASE aimed to assess handling skills, the test took place during, rather than at the end of, the course; therefore, it was perceived to have a positive impact on subsequent course moments by enhancing self-confidence. The assessment was found to provide more confidence and commitment during the learning process. An assessment performed between the first and the subsequent classes may also contribute to the levelling of the standard between students. Regarding whether the exam changed trainee perception, the opinions of the participants were divided: some considered that practical assessment did not have an impact, whereas others reported that it made a significant difference.

(5) OSLASE relevance and importance

Interviewees considered the OSLASE to be an essential component of the course, as it prepared them to perform practical work with animals, although it was also reported that students in general were motivated and would learn independently of the presence of a practical assessment. One participant highlighted its relevance from the perspective of responsibility regarding overseeing work with animals, in that not having a practical exam would require others, such as supervisors, to assess competence. Trainees in the introductory course considered the format to be adequate. Advanced course participants reported that the choice of the exam moment was adequate and contributed to the identification of specific improvement needs. Interestingly, one respondent discussed the relevance of a practical exam as a means to add authority to the argument that animal welfare is taken into account in the regulation of animal use, filling an existing gap in which their importance is undisputed but often not met in practice.

Discussion

The aim of this study was to explore the implementation of the OSLASE in LAS training, to evaluate the interrater agreement in the practical assessment and the acceptance of the exam by trainees.

The consistency of the agreement among raters provides valuable information regarding the quality of the assessment and its reliability. In particular, it reveals whether the raters agree on how to apply the instrument as a support for deciding on a candidate's performance. Two statistical methods that have been used in health sciences[17,23] were used to calculate the IRR: the percentage of agreement and Kw. These two statistical measures of agreement present advantages and limitations: although the percentage of agreement is easily calculated and directly interpretable, it does not consider whether the raters perform scoring randomly; in turn, the Kw formula corrects this limitation, but is based on assumptions regarding rater independence and other factors that are rarely met in real-life data, which may result in an extreme underestimation of the level of agreement.[17] The results of this study revealed that the majority of the items on all four scales exhibited moderate-to-good values of kappa, as well as acceptable percentages of agreement for all items but one. The Kw and percentage of agreement showed similar results, thus confirming a tendency on agreement, although some scale items were not fully concordant regarding the magnitude of accordance. Previous studies have reported moderate kappa values and excellent percentages of agreement between human and automated raters.[24]

Based on these results, there is room to enhance the IRR in the application of the OSLASE. One way of achieving this goal is to train the assessors further in how to apply the scale. Although the incorporation of

Table 3. Categories emerging from the focus group and representative quotes during the course: (1) choice of timing, (2) effect on trainee practical preparation, (3) experience and perception, (4) OSLASE impact on the trainee learning process, and (5) relevance and importance.

Phase	Category	Subcategory	Representative quotes
DURING	CHOICE OF TIMING	WHEN CONFIDENT – early	"I asked to [do the demonstration] during the first [class] because as I said, I was already more at ease with handling animals (…) I had done the other, more basic, course [with] the same practical component, therefore I felt comfortable to do it in the first class." [BS, ADVANCED COURSE]
		WHEN CONFIDENT – maximize training	"I, who had no experience in handling animals, felt that in the end of the first class was too early. It might even go well, let's say, but I don't know if I would have performed in the best possible way. Therefore I thought it (…), I would feel safer doing it in the next class, given that this was a possibility. And then at the end of the second class, I was more comfortable, I was more used to the technique and I think then it was easier too." [RCP, ADVANCED COURSE]
		NOT TOO LONG AFTER PRATICAL CLASSES	"(…) a month for me would be a little too much between the practical classes and the exam." [CP, INTRODUCTORY COURSE]
	EFFECT ON TRAINEE PRACTICAL PREPARATION	EXAM ENHANCES ENGAGEMENT	"It makes things more formal as regards the learning. I think we engage more in wanting to do it and to do it well because we know we will have to demonstrate later on this same knowledge that we learned." [JA, ADVANCED COURSE]
		EXAM DOES NOT AFFECT TRAINING	"I don't think it had any effect at all, in any case I was there to learn and certify that I had learned to work with mice, and therefore the demonstration was a validation in addition to my own, only that." [AC, ADVANCED COURSE] "For me no [it didn't affect how I prepared] because even some of the procedures were simulated in toys, not in live animals (…) I knew I had to do the procedures well because then I was going to do them in live animals anyway, so for me it didn't make a difference." [ML, INTRODUCTORY COURSE]
		EXAM HAS LITTLE INFLUENCE ON TRAINEES PREPARATION	"I was expecting there to be a practical exam because this makes sense for this course and therefore I didn't prepare differently, I prepared according to what I was expecting."[NM, INTRODUCTORY COURSE]
AFTER	EXPERIENCE AND PERCEPTION	EXAM INDUCED SOME ANXIETY	"(…) the preparation we had before allowed this demonstration to be made in a tranquil way." [PCR, ADVANCED]
		EXAM WAS NOT STRESSFUL	"(…) the preparation we had before allowed this demonstration to be made in a tranquil way." [PCR, ADVANCED] "(…) [the course having] few students [and] in turn the relation to the instructors was easy." [PLI, INTRODUCTORY COURSE]
	IMPACT ON TRAINEES' LEARNING PROCESS	POSITIVE IMPACT	"(…) now we have more confidence in ourselves and we have that internal dialogue that a competent person determined us as evaluated and that we perform handling well, and therefore we can proceed with our work." [RR, ADVANCED] "(…) the practical demonstration helps us completing the rest of the course with greater commitment." [CL, ADVANCED COURSE]
		NO IMPACT	"I think that the practical assessment in itself made no difference (…) it was really the theoretical assessment." [PC, INTRODUCTORY COURSE]
		ASSESSMENT AFFECTS THE EXPERIENCE WITH ANIMALS	"(…) in terms of experience and really dealing with the animals, handling them and doing the necessary protocols is really different with a theoretical assessment and a practical assessment." [NM, INTRODUCTORY COURSE]
		ASSESSMENT ESTABLISHES STANDARDS	"Perhaps if this evaluation would not be there, (…) in the next classes there may be members who would not feel so at ease with the activities, feel some difficulty, hence these two classes with evaluation ensure that we are more prepared for the subsequent classes which are more difficult in terms of procedures." [AC, ADVANCED COURSE]
	RELEVANCE AND IMPORTANCE	ASSESSEMENT IS CRUCIAL	" (…)if this is a course [that aims] to prepare us much for practising with laboratory animals the practical assessment cannot be missing." [NM, INTRODUCTORY COURSE]
		MIXED VIEW REGARDING THE EXAM RELEVANCE	" (…) if there were no evaluation would people learn? Yes. because I think all the people who were there were diligent, but taking into account the course it makes sense to have a [practical] exam." [BS, ADVANCED COURSE]
		IMPROVING EXISTING COMPETENCIES	"(…) only after the assessment did I effectively understand that there were some specific points that need improvement." [NM, INTRODUCTORY COURSE]
		IMPORTANT FOR EXTERNAL VALIDATION	"it gave me another tool for talking with people about the topic and be able to talk about one's own experience regarding the level of preparation and the level of training" by saying "look it's not quite that way; to do all this, we passed through training, we passed through verification" [PLI, INTRODUCTORY COURSE] "(…) having only a theoretical exam would place a much greater responsibility on the supervisor of the practical procedures right after the exam." [PLI, INTRODUCTORY COURSE]
		ADEQUATE EXAM	"(…) were assessed in the most basic, but also in the most common, animal manipulation and restraint, and the most common injections." [GLI, INTRODUCTORY COURSE] "(…) it was at the right moment and represented what we had learned before." [AC, ADVANCED COURSE] "(…) without knowing [how to handle animals well] it would be difficult to proceed to the rest." [SJ, ADVANCED COURSE]

continuous rater training into busy teaching schedules is challenging, it has been shown to considerably minimize disagreement, particularly for scale items that measure fine and very technical discriminations using various rating categories (as the scales used in this study).[25,26] In this study, performance was assessed over a three-year period, and despite being administered by assessors that were very familiar with the GRSs, the quality of the data collected was not monitored over time, to identify potential disagreements, which may have contributed to the limited IRR scores. Moreover, the scales were applied to a restricted number of performances in a controlled scenario (of training). The diversification of the context in which the instrument is applied, such as by including workplace assessments (of researchers performing the procedures as part of their work with animals) as part of a continuing professional development strategy, could enhance IRR.[27,28]

Going beyond the question of agreement, it is also relevant to consider the value of the OSLASE as a way to implement an assessment strategy that is able to evaluate the practical competence acquired by researchers during the training (for the advanced course format) or at the end of training (for the introductory format). This adapted OSCE type of exam introduced standardization and uniformization of the assessments for all trainees (independently of the course type). The existence of standardized instruments introduces predictability among the trainees, who know how they will be evaluated, and can obtain systematic feedback on their performance. The instruments applied at the stations, the GRSs, promote consistency, which guarantees that the manner in which the candidates' performance is evaluated is not dependent on who is performing the assessment.

Regarding the OSLASE acceptability, our findings revealed that the trainees accepted and recognized the validity of this type of structured practical assessment on the context of the LAS training course. The trainees took into consideration their confidence in their own capacity to decide on the timing to take the OSLASE, which seemed to be the key factor affecting the decision of when to perform the exam. It is noteworthy that trainees also highlighted another time issue: the time that is needed to develop and reach a level of proficiency to succeed in the assessment. The trainees who were forced to postpone the OSLASE because of COVID-19 pandemic restrictions reported frustration with the delay of the assessment, recognizing that the exam should be adjusted to the training timeline, as an excessive time interval between the end of classes and the exam was perceived as being problematic by them. The integration of the OSLASE in the course programme affected the preparation of the trainees: some respondents recognized that they felt more engaged and stimulated to invest on practical training. Furthermore, the trainees pointed out that assessments ensure that apprentices reach the expected practical learning outcomes. These findings are in line with research on the topic that supported the contention that assessment not only drives, but also has a positive impact on learning.[6,8] Although the OSLASE is a relatively recent introduction in this training, a few respondents stated that, for them, a formal practical assessment was assumed and expected in these courses. Assessment was not a stressful experience for the majority of the respondents, which may be explained by the manner in which the experienced tutors, also in the role of assessors, drove the process. In fact, the assessor role being assumed by the tutors can be regarded as a limitation of this study, as it can bias the way candidates faced of the OSLASE. However, the number of LAS tutors with expertise in teaching, supervising and assessing trainees is limited. The interviewees also identified the positive impact of the exam on the learning process, as reported previously in the literature.[9] The topic of how OSLASE helps the trainee to confirm the skills acquired emerged several times during the interviews, confirming the perception that assessment is an essential step in the consolidation and promotion of future practical learning.[29,30] The trainees also reported that the OSLASE establishes a standard and defines the skills that students are expected to develop in subsequent sessions. They also felt that demonstrations contributed to the identification of skills that should be improved and to the support of the role of those overseeing workplace training, which are crucial features introduced by the assessment in the context of these courses.

The strategy described in this paper was implemented in two course types, one of which is accredited by FELASA. Accreditation bodies are important in the efforts to support researcher competence development and maintenance during their professional life. Our experience suggests that having all trainees go through the same examination format mitigates the risk that trainees below the desired level of competence will complete the course, thus contributing to course credibility. However, the assessment and verification of the professional competence of individuals working with living animals is a process and a topic that goes beyond the responsibilities of the accreditation bodies or even course organizers. To guarantee that personnel performing procedures in animals are competent is also the responsibility of the institutions or companies at which the animals are used, the animal facilities and licensing authorities. FELASA and other accreditation bodies may play a fundamental role in supporting the change from a traditional paradigm of assessment by challenging the course organizers to include more extensive observational assessments in their courses. These approaches are relevant not only as summative, but also as formative assessments.

Conclusion

To the best of our knowledge, this was the first study to report the impact of the implementation of a practical assessment strategy in a LAS training context. In particular, this study addressed the feasibility of the OSLASE by exploring the interrater agreement and investigating the acceptance of the practical exam by trainees. The results showed that there is room for improvement of the interrater agreement, whereas the overall raters agreed on the decisions regarding the competence of the candidates. The trainees recognized that the OSLASE promoted and guided their learning, increased their engagement and motivation during practical sessions, and served as a confirmation that relevant skills had been learned. Overall, the introduction of the OSLASE allows the measurement of practical competence other than the assessment of cognitive knowledge using traditional evaluation methods.

Acknowledgements

The authors would like to give a special acknowledgement to all course participants who collaborated in this study. The authors also thank Dr Patricio Costa, from the University of Minho, for statistical assistance, and Anabela Nunes, from the i3S Communication Unit, for the support provided regarding the infographics of this paper.

Data availability

The data used in this study were collected from human participants who provided consent for the use of the data for research and publication purposes, but not for sharing as open data. The raw data collected for IRR calculation is available upon request to the corresponding author.

Declaration of conflicting interests

The author(s) declared no conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

ORCID iDs

Andreia Costa https://orcid.org/0000-0001-8285-0065
Sofia Lamas https://orcid.org/0000-0002-1665-1989
Maria S Gomes https://orcid.org/0000-0001-7657-2118
Manuel J Costa https://orcid.org/0000-0001-5255-4257
I Anna S Olsson https://orcid.org/0000-0002-4369-8723

References

1. European Commission. Guidance document. National Competent Authorities for the implementation of Directive 2010/63/EU on the protection of animals used for scientific purposes. A working document on the development of a common education and training framework to fulfil the requirements under the Directive. Brussels, http://ec.europa.eu/environment/chemicals/lab_animals/pdf/guidance/education_training/en.pdf (2014, accessed 13 April 2020).
2. European Commission. Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. *Official Journal of the European Union*. Brussels: European Commission, 2010, L276/33.
3. NC3Rs (National Centre for the Replacement RRoAiR. *The 3Rs*, https://nc3rs.org.uk/the-3rs (2022, accessed 10 january 2022).
4. Duffee N, Nevalainen T and Hau J. Education and training. In: Hau J and Schapiro SJ (eds) *Handbook of laboratory animal science: Essential principles and practices* 4th ed. Boca Raton, FL: CRC Press, 2021, p.12.
5. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990; 65 (9 Suppl.): S63–S67.
6. Wass V, van der Vleuten C, Shatzer J, et al. Assessment of clinical competence. *Lancet* 2001; 357: 945–949.
7. Bashook PG. Best practices for assessing competence and performance of the behavioral health workforce. *Adm Policy Ment Health* 2005; 32: 563–592.
8. Tabish SA. Assessment methods in medical education. *Int J Health Sci (Qassim)* 2008; 2: 3–7.
9. Preston R, Gratani M, Owens K, et al. Exploring the impact of assessment on medical students' learning. *Assess Eval High Educ* 2020; 45: 109–124.
10. Hecker KG, Norris J and Coe JB. Workplace-based assessment in a primary-care setting. *J Vet Med Educ* 2012; 39: 229–240.
11. Costa A, Lamas S, Costa MJ, et al. The assessment of researchers' competence in experimental procedures with laboratory animals: A three-step methodology to develop a global rating scale. *Lab Anim* 2021; 55: 463–471.
12. Epstein RM and Hundert EM. Defining and assessing professional competence. *JAMA* 2002; 287: 226–235.
13. Laboratory Animal Science Association (LASA). Direct Observation of Practical/Procedural Skills (DOPS), https://www.lasa.co.uk/dops/ (accessed 1 April 2022).
14. Education and Training Platform for Laboratory Animal Science. DOPS and SOP, https://etplas.eu/education/dops-and-sop/ (accessed 1 April 2022).
15. Costa A, Lamas S, Costa MJ, et al. Global rating scales to assess the competence of researchers in handling and performing substance administrations in laboratory rodents, https://repositorio-aberto.up.pt/handle/10216/133419 (2021, accessed 19 April 2021).
16. Harden RM, Stevenson M, Downie WW, et al. Assessment of clinical competence using objective structured examination. *Br Med J* 1975; 1: 447–451.
17. McHugh ML. Interrater reliability: The kappa statistic. *Biochem Med (Zagreb)* 2012; 22: 276–282.
18. Graham MJ, Milanowski AT and Miller JB (eds). Measuring and promoting inter-rater agreement of teacher and principal performance ratings, https://files.eric.ed.gov/fulltext/ED532068.pdf (2012, accessed 19 April 2022].
19. Fleiss JL, Cohen J and Everitt BS. Large sample standard errors of kappa and weighted kappa. *Psychol Bull* 1969; 72: 323–327.
20. Cohen J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull* 1968; 70: 213–220.
21. Chmura Kraemer H, Periyakoil VS and Noda A. Kappa coefficients in medical research. *Stat Med* 2002; 21: 2109–2129.
22. Altman DG (ed.) *Practical statistics for medical research*. London: Chapman and Hall, 1991.
23. Hripcsak G and Heitjan DF. Measuring agreement in medical informatics reliability studies. *J Biomed Inform* 2002; 35: 99–110.
24. Simundic AM, Nikolac N, Ivankovic V, et al. Comparison of visual vs. automated detection of lipemic, icteric and hemolyzed specimens: Can we rely on a human eye? *Clin Chem Lab Med* 2009; 47: 1361–1365.
25. Kardong-Edgren S, Oermann MH, Rizzolo MA, et al. Establishing inter- and intrarater reliability for high-stakes testing using simulation. *Nurs Educ Perspect* 2017; 38: 63–68.
26. Holland AE, Tiffany J, Blazovich L, et al. The effect of evaluator training on inter-and intrarater reliability in high-stakes assessment in simulation. *Nurs Educ Perspect* 2020; 41: 222–228.
27. Crossley J, Humphris G and Jolly B. Assessing health professionals. *Med Educ* 2002; 36: 800–804.
28. Van Dalen J, Prince CJAH, Scherpbier AJJA, et al. Evaluating communication skills. *Adv Health Sci Educ Theory Pract* 1998; 3: 187–195.

29. Bidonde J, Meneses-Echavez JF, Asare B, et al. Developing a tool to assess the skills to perform a health technology assessment. *BMC Med Res Methodol* 2022; 22: 78.
30. Heeneman S, Oudkerk Pool A, Schuwirth LW, et al. The impact of programmatic assessment on student learning: Theory versus practice. *Med Educ* 2015; 49: 487–498.

Examen structuré objectif de science animale de laboratoire (OSASSE) pour assurer la compétence professionnelle des chercheurs en SAL

Résumé

L'évaluation de la compétence du personnel travaillant avec des animaux de laboratoire présente actuellement un défi. La directive 2010/63/UE établit que le personnel doit avoir démontré sa compétence avant d'effectuer un travail non supervisé avec des animaux vivants. Il existe toutefois un manque de recherche sur l'éducation et la formation en science animale de laboratoire (SAL), et l'établissement de stratégies d'évaluation pour confirmer la compétence des chercheurs n'est pas encore abordé.
Dans cette étude, nous avons analysé la mise en œuvre d'une stratégie d'évaluation pratique sur 3 années consécutives (2018–2021) à l'aide de l'examen scientifique structuré objectif de science animale de laboratoire (OSLASE) développé précédemment par nous pour évaluer la compétence professionnelle. La concordance inter-juge (CIJ) était déterminée en fonction de l'évaluation par les évaluateurs de la performance des candidats à différentes stations de l'OSASSE en utilisant le Kappa pondéré (Kp) et le pourcentage d'accord. Des entrevues de groupe de discussion ont été menées pour avoir accès à l'acceptabilité des stagiaires concernant l'OSASSE.
Le Kp s'avérait « modéré à bon » pour la majorité des éléments de l'échelle ($0{,}79 \pm 0{,}20 \leq \mathrm{Kp} \geq 0{,}45 \pm 0{,}13$). Les pourcentages d'accord étaient également acceptables ($\geq$75%) pour tous les éléments de l'échelle, à l'exception d'un. Les stagiaires ont indiqué que l'OSASSE avait eu un impact positif sur leur engagement au cours de la formation pratique et qu'il avait clarifié les normes établies concernant leur performance et es compétences nécessitant une amélioration. Ces résultats préliminaires illustrent comment des stratégies d'évaluation, telles que l'OSASSE, peuvent être mises en œuvre d'une manière utile tant pour les évaluateurs que pour les stagiaires.
Eine objektive strukturierte Prüfung für Labortierkunde (Objective Structured Laboratory Animal Science Examination, OSLASE) zur Sicherstellung der fachlichen Kompetenz von Forschern in der Labortierkunde (LAS)
Abstract

Eine objektive strukturierte Prüfung für Labortierkunde (Objective Structured Laboratory Animal Science Examination, OSLASE) zur Sicherstellung der fachlichen Kompetenz von Forschern in der Labortierkunde (LAS)

Abstract

Die Bewertung der Kompetenz von Personal, das mit Labortieren arbeitet, ist derzeit problematisch. Die Richtlinie 2010/63/EU legt fest, dass das Personal entsprechende Kompetenz nachgewiesen haben muss, bevor es unbeaufsichtigte Arbeiten mit lebenden Tieren durchführt. Es mangelt jedoch an Forschungsarbeiten zu Aus- und Weiterbildung im Bereich der Labortierkunde (LAS), und die Festlegung von Bewertungsstrategien zur Bestätigung der Kompetenz von Forschern bleibt weitgehend ungeklärt. In dieser Studie analysierten wir die Umsetzung einer praktischen Bewertungsstrategie über drei aufeinanderfolgende Jahre (2018–2021) unter Verwendung der von uns zuvor entwickelten Objective Structured Laboratory Animal Science Exam (OSLASE) zur Bewertung der beruflichen Kompetenz. Die Interrater-Reliabilität (IRR) wurde auf der Grundlage der Bewertung der Leistung der Kandidaten an den verschiedenen OSLASE-Stationen durch die Prüfer mit Hilfe von Weighted Kappa (Kw) und dem Prozentsatz der Übereinstimmung ermittelt. Es wurden Fokusgruppeninterviews durchgeführt, um die Akzeptanz der OSLASE bei den Auszubildenden zu ermitteln. Für die meisten Items der Skalen ergab sich ein moderater bis guter Kw-Wert ($0{,}79 \pm 0{,}20 \leq \mathrm{Kw} \geq 0{,}45 \pm 0{,}13$). Auch die prozentuale Übereinstimmung war bei allen Skalenitems, außer einem, akzeptabel ($\geq$75%). Die Kandidaten berichteten, dass sich die OSLASE positiv auf ihr Engagement während der praktischen Ausbildung auswirkte und dass sie die für ihre Leistung festgelegten Normen und die zu verbessernden Fähigkeiten klar verdeutlichte. Diese vorläufigen Ergebnisse veranschaulichen, wie Bewertungsstrategien wie OSLASE in einer Weise umgesetzt werden können, die sowohl für Prüfer als auch für Auszubildende nützlich ist.

Un estudio científico objetivo estructurado sobre los animales de laboratorio (OSLASE) para garantizar la competencia profesional de los investigadores en LAS

Resumen

Un estudio científico objetivo estructurado sobre los animales de laboratorio (OSLASE) para garantizar la competencia profesional de los investigadores en LAS

Resumen

La evaluación de la competencia del personal que trabaja con animales de laboratorio es actualmente un verdadero reto. La Directiva 2010/63/UE indica que el personal debe haber demostrado su competencia antes de realizar trabajos no supervisados con animales vivos. No obstante, existe una falta de investigación sobre la educación y la formación en la ciencia de los animales de laboratorio (LAS), y el establecimiento de estrategias de evaluación para confirmar la competencia de los investigadores continua, en gran medida, sin abordarse.

En este estudio analizamos la implementación de una estrategia de evaluación práctica durante 3 años consecutivos (2018–2021) utilizando el estudio científico objetivo estructurado sobre los animales de laboratorio (OSLASE) desarrollado previamente por nosotros para evaluar la competencia profesional. Se determinó la fiabilidad entre evaluadores (IRR) a partir de la calificación de los evaluadores sobre el rendimiento de los candidatos en diferentes estaciones del OSLASE utilizando Weighed Kappa (Kw) y el porcentaje de acuerdo. Se realizaron entrevistas en grupos de discusión para acceder a la aceptabilidad de los aprendices en relación con el OSLASE.

Hubo un Kw de moderado a bueno para la mayoría de los ítems de las escalas ($0{,}79 \pm 0{,}20 \leq Kw \geq 0{,}45 \pm 0{,}13$). Los porcentajes de acuerdo también fueron aceptables ($\geq$75%) para todos los ítems de la escala salvo uno. Los estudiantes informaron de que el OSLASE tuvo un impacto positivo en su compromiso durante la formación práctica y que aclaró los estándares establecidos para su rendimiento y las habilidades que requerían ser mejoradas. Estos resultados preliminares ilustran cómo las estrategias de evaluación, como el OSLASE, pueden ser implementadas de manera que sean útiles tanto para los evaluadores como para los estudiantes.