



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 João Costa⁴, I. Anna S. Olsson^{1,2}

¹i3S - Instituto de Investigação e Inovação em Saúde, University of Porto, Portugal

²IBMC - Institute for Molecular and Cell Biology, University of Porto, Portugal

³ICBAS - Abel Salazar Institute of Biomedical Sciences, University of Porto, Portugal

⁴ICVS - Life and Health Sciences Research Institute, School of Medicine, University of Minho, Braga, Portugal

* Contact author: afcosta@ibmc.up.pt

Published in *Laboratory Animals* / Article first published online: May 29, 2021

<https://doi.org/10.1177/00236772211017767>

Keywords:

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

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

Version: Postprint (identical content as published paper) This is a self-archived document from i3S – Instituto de Investigação e Inovação em Saúde in the University of Porto Open Repository For Open Access to more of our publications, please visit <http://repositorio-aberto.up.pt/>



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 programs 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 behavioral 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.

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 programs, 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 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 endeavor, 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 behavioral 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: i. theoretical and technical knowledge (e.g. of maximum volumes of solution per administration, the selection of an appropriate route and/or site);

**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

Version: Postprint (identical content as published paper) This is a self-archived document from i3S – Instituto de Investigação e Inovação em Saúde in the University of Porto Open Repository For Open Access to more of our publications, please visit <http://repositorio-aberto.up.pt/>

ii. technical ability (e.g. in animal restraint or needle/syringe manipulation); and iii. attitudinal ability (including motivation, professionalism, values, and 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 behaviors, whether in controlled or 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 (i) knows and (ii) knows how (both cognition). Then there is (iii) shows how and (iv) does (performative and behavioral). At cognition levels (i) and (ii) competence is usually assessed through essays or written tests designed with close-ended or selected response questions (see 7, 8). At higher levels (iii) and (iv), behavior 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 – e.g. 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 to related fields such as medicine and veterinary medicine, the literature on assessment methodology for laboratory animal science 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 (LASA) [21]. The Education and Training Platform for Laboratory Animal Science (ETPLAS) have been working through their working groups with aim to develop Directly Observed Practical Skills (DOPS) 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

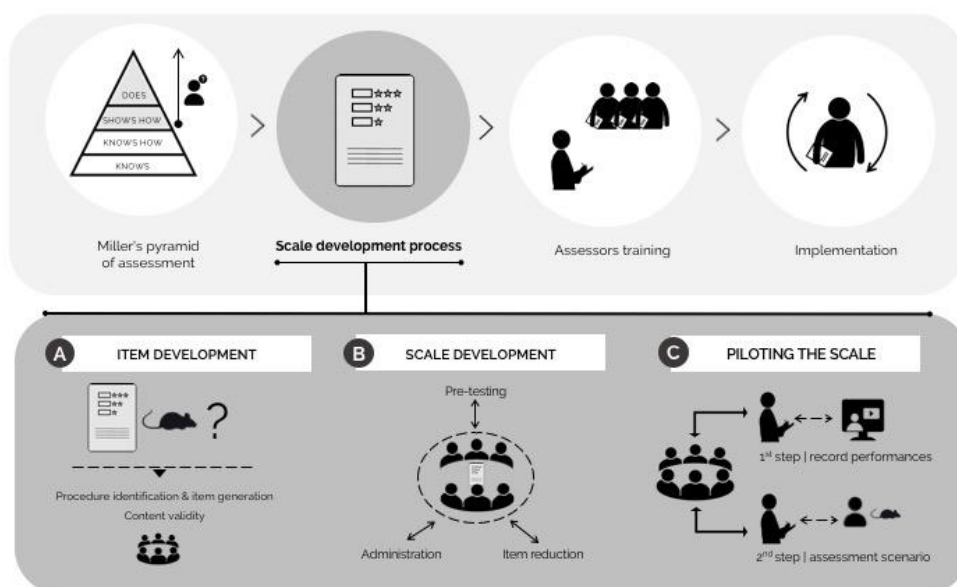


Figure 1. Scheme illustrating the operationalization of Miller's pyramid of assessment to evaluate professional competence in laboratory rodent experimentation at the performative and behavioural levels. The figure highlights the most representative stages of scale development: A. Item Development, B.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

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 objectives²⁴) 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 to 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 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.

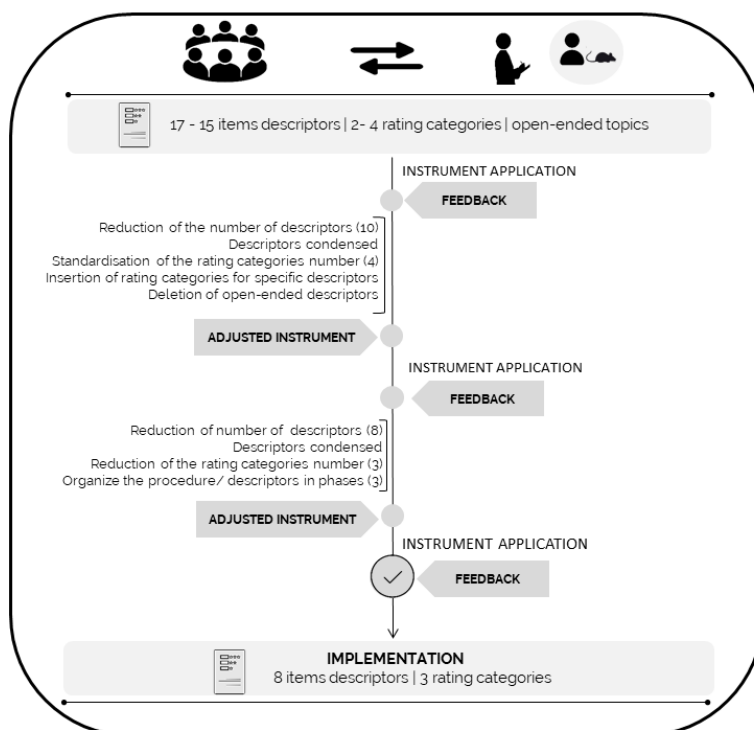


Figure 2. Steps in the adjustment of the instrument mainly resulting from phase C. Piloting of the Scale. The definitions of educational terms are provided by the National Council on Measurement in Education: <https://www.ncme.org/resources/glossary>.

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 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).

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, Ponnampereuma 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 bi methodologies. *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. Jennings 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.