Received: 18 SET 2025 | Accepted: 08 OUT 2025 | **DOI:** https://doi.org/10.57849/ulisboa.fm.jscml.0000031.2025

JORNAL DA SOCIEDADE DAS CIÊNCIAS MÉDICAS DE LISBOA

Assessment methods in Undergraduate Medical Education: brief considerations

Part of the content presented in this article is drawn from the Pedagogical Report of the Aggregation Examination in Medicine (Neurology), conducted at the Faculty of Medicine, University of Lisbon (May 2025).

Patrícia Canhão [1]

[1] Centro de Estudos Egas Moniz, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal

✉ **Corresponding author:**
Patrícia Canhão
Email: *pcanhao@fm.ul.pt*

ABSTRACT: Student assessment is an integral part of a medical curriculum. One of the main objectives is to verify whether students meet the standards of knowledge and performance that allow them to progress to subsequent levels of learning. To achieve this aim, assessment methods must be developed in accordance with the learning objectives defined for the stage at which students are at the time of assessment.

For the assessment of a medical student to be complete, it is necessary to use various methods to evaluate different crucial competencies: knowledge, clinical skills, and attitudes. This paper reviews the best known or most frequently used tools for each of these competencies, such as written or oral assessments, multiple-choice tests with particular emphasis for single best answer (SBA) questions, direct observation in learning environments, objective structured clinical examinations (OSCE), and the use of portfolios. The use of simulation is addressed in various assessment contexts. A brief description is provided of global assessment scales, tools that confer objectivity and validity on assessments and improve reproducibility among assessors. The most important properties that an assessment must fulfil in order to be useful are described. Particular importance is given to psychometric criteria such as validity, reliability, feasibility and acceptability.

Whether for formative or summative purposes, student assessment is an essential step in the learning process. The analysis of the results of the various assessments should guide medical schools in optimising teaching and learning processes. For students, it should contribute to their self-learning, so that they can play an active role in the process of their own improvement.

KEYWORDS: Medical education; Assessment; Evaluation; Knowledge and Skills

INTRODUCTION

Student assessment is an integral part of a medical curriculum. One of the fundamental objectives of assessment is to verify whether students meet the standards of knowledge and performance that allow them to progress to subsequent levels of learning. Another relevant attribute is that, through the exposure of learning objectives, students identify the areas of the course that are considered important, which motivates study and improvement in essential areas. On the other hand, following the assessment results, students receive information about their performance, weaknesses, and ways to improve. Additionally, and of interest to faculty, it can measure the effectiveness of teaching and identify weaknesses in the curriculum or its implementation.

Until around 1950, knowledge and clinical skills were essentially assessed through written or oral exams. The questions on written exams were often open-ended, and were often corrected in a very subjective manner. Oral examinations typically involved an interview and observation of a patient, followed by a discussion with the assessors, who scored the student's performance and questioned them on aspects considered relevant. [1] This type of assessment had several weaknesses, related to subjectivity in the evaluation of the exams, heterogeneity of the evaluators and the patients, and the long time required for its implementation.

From the 1950s to the present day, we have witnessed several changes in the way medical education is assessed. [2] Three types of factors have contributed to this. The first factor was the development of new assessment methods, with the aim of measuring the different areas of competence of a future doctor. Examples of these new methods include the introduction of multiple-choice tests, the use of structured clinical examination stations known as OSCE (Objective Structured Clinical Examination), the use of simulation environments with actors (standardized patients), and other techniques to assess non-cognitive skills, such as communication or professionalism.[3] A second factor was the increasingly widespread adoption of computers as an integral part of assessment. In addition to their central role in the construction and administration of multiple-choice tests, computers provide important support in the process of obtaining scores and analysing assessment metrics. Furthermore, computers and new digital technologies have been increasingly explored for their potential to create simulation environments, generating realistic clinical scenarios in which students can be assessed without subjecting real patients to this type of evaluation. The third factor that has contributed to the change in the assessment relates to the development of research in Medical Education, with increasing research on the psychometric principles that assessments must comply with, and the identification of various types of faults in assessment processes and proposals for their reduction. [2,4]

This paper reviews several methods of assessing undergraduate students for the different skills that a doctor must have (knowledge, skills, and attitudes), and the most important psychometric properties that these instruments must fulfil in order to be fair and accurate.

1. ASSESSMENT METHODS

> "*No single assessment method can provide all the data required for judgment of anything so complex as the delivery of professional services by a successful physician*"
> (Miller, 1990) [5]

For an assessment to fulfil its function properly, it is first necessary to know and define what is to be assessed. Thus, even before defining the assessment methods and instruments, it is important to define the learning objectives: knowledge, clinical skills, procedures, and attitudes that students must demonstrate in order to be considered competent for the curricular level at which they are studying. The selection of methods and construction of assessment tools should focus on the assessment of these endpoints to be achieved. The objectives must be clear, such as what type of task the student should be able to perform, the conditions under which that task should be required, and the level of performance considered acceptable. The definition of objectives must be precise and specific to ensure the quality of the assessment. [6]

Learning objectives are broad and complex, and it is difficult for a single tool to analyse all the areas in which students should be evaluated. Therefore, it is mandatory to use several assessment methods in order to achieve an accurate and comprehensive assessment.

In brief, there are three types of competencies that should be assessed: knowledge, skills, and attitudes. Knowledge refers to theoretical information of the curriculum that must be learned and understood;

skills refer to the ability to perform tasks, such as taking a medical history, performing a physical examination or a technical procedure; attitudes refer to professional qualities that are essential for a future doctor, such as communication skills, sense of responsibility, ethics, and empathy. Once the learning endpoints for these several competencies have been defined, different forms of assessment can be developed for the various fields to be tested. Importantly, to be more successful in achieving these endpoints, students should be informed of the learning objectives.

There is currently a wide variety of assessment tools: essays, projects, reports, patient problem management, short-answer questions, multiple-choice questions, rapid clinical case assessment, OSCE, checklists, rating scales, portfolios.[7] These instruments can be used in a complementary way, with some mainly assessing knowledge and cognitive processes, and others assessing technical skills or attitudes. Below, we describe the best known or most frequently used ones.

1.1. Tools for assessing knowledge

Written assessments are likely the most commonly used methods in medical education. They are relatively easy to conduct in terms of human resources and costs. They are used to assess knowledge, clinical reasoning, problem-solving, and decision-making skills in specific clinical scenarios. However, they are not suitable for assessing other curricular competencies, such as clinical skills, attitudes, or professionalism. Oral assessments can also be used to test knowledge and clinical reasoning, but they are rarely used because they are time-consuming, somewhat subjective, assess only limited aspects of the curriculum, and are subject to various biases, which can lead to inaccurate or poor reliable ratings.

There are several types of written assessments. One of these is the use of open-ended questions, in which students must produce a spontaneous response. These include short-answer questions, in which students must respond with two to three words, or questions that require longer answers, or even broader responses such as essays. Alternatively, closed questions can be used. In this type of questions, students are asked to select answers from a list of options. Examples of closed questions are: true/false questions, in which students must categorize the answer options as true or false; multiple-choice questions, also known as single best answer (SBA) questions, in which the student must select the most correct option from a group of alternatives; multiple true-false questions, in which the candidate can choose more than one option.[8]

It has increasingly been advised to use written assessments with closed questions, as they are logistically easier to apply and rate. Within these, it has been proposed to abandon true or false questions and give preference to single-best-answer (SBA) questions.[9] The SBA question is composed of three parts: the stem (e.g., a clinical case presentation), the lead-in question, and the options, in which one is "the most correct" among three to five distractors.[9] When properly constructed, multiple-choice questions allow for the assessment of not only knowledge of specific content, but also other characteristics such as the ability to interpret, synthesize, and judge in concrete clinical scenarios.[10] In other words, they entail the use of complex cognitive processes that replicate clinical practice.

SBA questions take time and effort to be constructed in order to allow for a valid and fair assessment of students. [11,12] This type of assessment, however, has the great advantage of being able to examine a large number of students in a short period of time, i.e., they meet feasibility and cost-effectiveness requirements. For example, they are the type of questions currently used in Portugal in the National Access Test (PNA) required for doctors to access a specialist training position. (Despacho nº. 4412/2018, de 04 de maio).

Regardless of the method used to assess knowledge, the overall quality of an assessment test depends on the quality of the individual questions (items). Several features should be considered when evaluating the quality of items: the relevance of the items; the alignment between the curriculum objectives and the content of the items; the level of difficulty; the ability to discriminate between students based on their level of knowledge; and student feedback.

1.2. Tools for assessing skills and attitudes

> *"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"*
> (Klass, 2000) [13]

The most appropriate tool for assessing the performance of students with patients should ideally be applied in real situations with supervision and obser-

vation of performance over a specific period of time. It can be carried out in learning environments, where it has adequate apparent validity. However, it may entail several limitations, such as a lack of standardization among assessors, the influence of the clinical contexts in which it occurs, the restriction of the skills being evaluated, and the number of hours required to obtain valid assessments. This format of direct student observation is commonly carried out in continuous assessments, using global performance grids or scales (see below). It can focus on clinical skills, communication skills, teamwork, presentation skills, work habits, and organization. Theoretically, there are several possible types of evaluators: teachers, hospital staff, patients, colleagues, and self-assessment. These observations can be recorded in reports, checklists, or rating scales.[14,15]

When this longitudinal assessment in a real-world setting is not feasible, other formats of direct and standardized observation of students can be used. Examples include long case exercises or mini-clinical evaluation exercises (mini-CEX). [16,17] In these assessment formats, students can be observed while performing certain tasks, such as taking a medical history and performing a physical examination, for a period of 10 to 20 minutes. After that, the students are expected to present a diagnosis and treatment plan, and faculty members evaluate, discuss, and provide educational feedback. This direct observation of students can be combined with exercises that students perform after contact with patients, such as oral case presentations or written exercises that assess reasoning. Videos of encounters with patients can also be used, with the great advantage of evaluating and providing feedback on students' performance in clinical interactions. These approaches are suitable for assessing communication and technical skills, but they are selective in the skills and attitudes they analyse, require time, and are not feasible for evaluating a large number of students.

Due to various difficulties and limitations in the use of real patients for assessment, simulated patients (standardized patients) are increasingly being used. These are actors trained to assume the role of a patient, giving a consistent history, in a specific style of communication and physical findings. Interactions with standardized patients should be adapted to meet specific educational objectives. Standardized patients are often integrated into Objective Structured Clinical Examinations (OSCE). This type of assessment was introduced by Harden and Gleeson, and consists of a series of timed stations, each focused on a different task. [18] Each student must rotate through each station for about 5-10 minutes, where they are challenged with a clinical scenario in which they must demonstrate specific skills and attitudes. A hallmark of the OSCE is that each student must perform the same sequence of tasks and be rated using a standardized scheme. Faculty members use a checklist of specific skills and attitudes, or an overall rating form to assess student performance. [19,20]

Over the years, OSCE have been adapted to integrate assessments of multiple skills: communication, physical examination, interpretation of diagnostic tests, technical skills. The use of simulation models has broadened the possibilities for assessing some more complex skills. [21]

One issue to consider regarding the validity of the OSCE is its ability to assess performance-based skills. The greater the number of stations, the greater the reliability of the assessment. An OSCE with 25 stations over 8 hours may result in excellent reliability, but it is not realistic for most universities, raising issues of feasibility. For this type of assessment to be highly reliable, it has been suggested that it should comprise at least 10 stations. [22]

There are certain procedures and techniques that are difficult or even impossible to test on real or simulated patients. Examples of these include collecting venous or arterial blood, emergency procedures, and surgical techniques. To evaluate these practical procedures, simulations involving sophisticated models are increasingly being used. In fact, simulation is increasingly seen as an important learning aid and can be useful in assessing knowledge, technical procedures, and even teamwork skills.

It should be noted that there is no evidence that a student's performance in an OSCE or simulation environment is predictive of their performance with a real patient. In fact, it has been shown that a student's performance in real life is worse than in simulated environments. [23] Therefore, the use of other methods to assess student's skills and attitudes in the real work environment is justified. The use of diaries, portfolios, or logbooks can be additional tools that reveal what the student has seen or done. They do not replace direct observation of the student, but they can be quite informative.

1.3. Portfolios

A portfolio is a collection of a student's work, which includes documentation and reflections on specific areas of competency, and helps to demonstrate their progress and expertise. It can cover different learning materials, such as records of patients, tasks performed or observed, learning about diseases, literature research, and projects for self-improvement.

Portfolios are useful for demonstrating the acquisition of knowledge, skills, activities, attitudes, and progress over time. They often include self-assessments and learning plans. Although they are most commonly used in formative assessment, they can also be used for summative assessments.

One of the most relevant characteristics of the portfolio is that it leads to self-reflection, revisiting the learning experience to extract the maximum amount of teaching from it. This self-reflection can be extremely useful for promoting self-learning, personal and professional development, and refining practice. [24]

The process of assessing a student through a portfolio comprises several steps: analysis of the learning collection; analysis of reflection on learning, steps that the student took to complement it; assessment of what the student actually learned and accomplished, in relation to the learning objectives; discussion with the student about the evidence of learning; and finally, assessment rating.

There are potential advantages to using portfolios as summative assessment: students demonstrate the scope and depth of the content they have observed; they do so continuously over time; they demonstrate professionalism in achieving learning endpoints; they indicate self-reflection and self-learning; and they allow student's expression of creativity and identity. However, there are issues limiting their use as summative assessment: irregular collaboration by students in completing portfolios; variability among evaluators; and time required for proper analysis. It is therefore essential to decide on the purpose of using portfolios: as a tool for formative or summative assessment. If it is decided that it should be a summative assessment tool, specific objectives should be defined for assessing students: endpoints to be achieved; definition of portfolio content to verify the achievement of learning objectives; activity scoring system; criteria and scales for standardized grading by examiners.[25]

2. GLOBAL ASSESSMENT SCALES

These scales are extremely useful for assessing skills and attitudes that cannot be measured in written tests. They are used to assess clinical skills, technical procedures, communication or other skills related to professionalism. They can be used while observing the student's performance or by recalling their interventions, for example, in practical classes. One of the objectives of using rating scales is to reduce the heterogeneity of classification between different assessors.

Two types of scales are commonly used for quantifying student's performance: rating scales anchored by behavior and Likert scales. Rating scales anchored by behavior were introduced in 1963 by Smith and Kendall. [26] They are constructed based on expected behaviors considered to be "anchors" in the domain under study. Assessors evaluate students by scoring their behaviors on a scale from the lowest to the highest value. Establishing specific behaviors for classification gives the classification a higher degree of accuracy in relation to performance. The number of points on the scale is determined by the number of "anchors" and the evaluator's ability to discriminate between the points on the scale. In some cases, the scale can be grouped into three or more anchors, with more than one point assigned to each, which allows for greater flexibility in the evaluation. Likert scales were developed in 1932. [27] They can take various forms: evaluation (poor-excellent), frequency (never-always), agreement (agree-disagree). A numerical value is assigned to each category. Points on the scale are assigned from the lowest to the highest value. Examiners score skills and attitudes on a continuous scale, for example from "poor" to "excellent."

These scales are easy to develop and apply. However, some flaws can interfere with the validity or reproducibility of the scales. [28] One of these is the "halo effect," in which the examiner is influenced by some characteristics of the student, and consistently scores different aspects of the assessment at the same point on the scale. Another error is the "central tendency," in which the examiner tends to avoid extreme scores. Another problem that can interfere with the validity and reliability of the assessment depends on the examiner, who may be excessively or insufficiently demanding, depending on their expectations of the performance that the examinee should demonstrate. One such error is the "contrast error," where examiners who are exceptionally competent in an area tend to score students lower than other assessors. One way to analyse these

kind of discrepancies is through inter-rater agreement analysis. An effective way to reduce potential errors is through assessor training. Limitations are also minimized through the participation of multiple examiners, for example in OSCE or continuous assessments.

Global assessment scales can be supplemented with checklists. In certain circumstances, it is appropriate to record the performance of specific procedures or tasks. In such cases, a checklist can be used for the various behaviors that the student must exhibit when solving a specific problem. In these circumstances, the assessor scores whether or not the student completed the procedures or tasks that were required. This type of assessment is common in OSCE and continuous assessments. They are objective, quick to carry out, and can be standardized for assessing all students. One drawback is that they do not provide information on the quality of the task performance, which means that they should only be used to evaluate procedures that do not require an assessment of the quality with which they are performed.

3. SELF-ASSESSMENT

"*Self-assessment may be regarded as a skill and, as such, needs to be developed*"
(Falchikov & Boud, 1989) [29]

Lifelong learning requires not only the ability to work independently, but also to assess one's own performance and progress. In medical education, self-assessment refers to the involvement of students in making decisions about their learning, particularly regarding the acquisition of their performance and achievements. Self-assessment is formative in that it contributes to the learning process and helps students to focus on specific areas that require improvement. Rather than simply receiving assessments from external sources, students are encouraged to become aware of their own development and progress.

The self-assessment process may include aspects such as performance review and identification of strengths and weaknesses. Any form of evaluation can be used for self-assessment, provided that the student is confronted with the "gold standard" criteria against which they can compare themselves. It can result from multiple-choice tests and OSCE, once they have access to feedback. Portfolios or logbooks are very comprehensive methods of self-assessment, requiring a high level of self-assessment of several types of performance. There is not always an overlap between the level of performance estimated by students and that identified by the Faculty, which is a very interesting area of research, particularly to understand the reasons for these discrepancies and to what extent the reasons for these discrepancies can help to identify specific pedagogical interventions.

4. PROPERTIES OF THE ASSESSMENTS

Van Der Vleuten described five criteria for determining the usefulness of a specific assessment method: validity (whether the assessment measures what it claims to measure), reliability or reproducibility (the degree to which the measurement is accurate and reproducible), impact on future learning and practice, acceptability (to students and institution), and costs (to the student and institution).[30] For an assessment tool to be useful and fair in evaluating student performance, it must meet these criteria as much as possible.

4.1. Validity

Regardless of the assessment tool, it must be valid for measuring the task for which it was designed. There are several types of validity: content validity, concurrent validity, predictive validity, construct validity, and face validity. [7,30] Some of these forms of validity are more decisive in the context of student assessment. Perhaps the most important in undergraduate assessment is content validity. Content validity refers to the accuracy with which a test assesses content that is included in the learning objectives. In addition to ensuring the validity of individual items (questions about learning objectives), it must also ensure "sampling validity," which means that a test should include a representative sample of the different learning objectives.

Concurrent validity involves demonstrating that the scores on one test correlate with the scores on another test that was constructed previously and administered after the new test. Predictive validity refers to the ability of a test to predict future performance. Construct validity refers to the degree to which a test assesses the underlying theoretical construct it is supposed to measure. Apparent validity is determined by the feedback received from students on whether the test seems valid to them.

There are several factors that reduce validity and should be identified: vague instructions to students, overly complicated vocabulary, a small number

of questions (poor sampling validity), insufficient time for assessment, items that are too easy or too difficult in relation to the learning objectives, and topics that are inappropriate for the learning objectives.

4.2. Reliability

Reliability, or reproducibility, is the degree to which the instrument measures what it is supposed to measure in an accurate, consistent, and reproducible manner. There are several ways to test reliability. [7,30] The internal consistency of a test can be calculated by dividing the test into two parts and analysing the agreement between the two sets of scores. This has the advantage of being calculable with only one test, and is more reliable when the test has a larger number of items. There are several statistical techniques for calculating the internal consistency of a test, the best known being Cronbach's α and KR20. Another form of reliability refers to intra- or inter-examiner reliability. This type of concern is particularly relevant when the assessment requires some degree of subjectivity. In such cases, it may be appropriate to calculate the agreement between evaluators.

There are other ways to check reliability, but they need a number of different tests. To see how consistent a test is over time (test-retest), the same test is given to the same group of students at two different times. Another way to test reliability is by using essentially similar tests, in which modifications are made to the initial test without changing its essential objectives, and applying the second test to the same students to assess the consistency of the scores between the two tests.

There are some factors that increase reliability: the size of the assessment and objectivity in scoring. There are errors that reduce reliability, such as mistakes in scoring, generalization, or bias in student assessment.

4.3. Feasibility

The development of an assessment should take into account the availability of human resources, space, time, and costs. Efforts should be made to find the greatest number of assessments that can evaluate the different components of competence: knowledge, skills, and attitudes. Several questions should be asked to support the various selections: how long does it take to develop the instrument? How long does it take to be completed? How will the scores be interpreted? Is it practical to implement? Is it beneficial for students?

5. FORMATIVE OR SUMMATIVE ASSESSMENT

The type of assessment to be implemented will depend on two types of approach: formative and summative. Formative assessment focuses on guiding students in their learning process, providing reassurance, and promoting self-reflection. This type of assessment, which is highly valued by students, is extremely important provided that it is followed by an intervention program. Summative assessment allows for a general judgment to be made about competence and aptitude for practice. In other words, students are assessed to decide whether they have the minimum qualifications to proceed to the next level. A distinction must be made between assessments that are only suitable for formative use and those that have sufficient psychometric rigor for summative use.

If the aim is to identify students' qualifications, ensuring that they are qualified to proceed to the next level, it is advisable to use summative assessment methods. Summative assessments may not provide feedback. However, summative assessment can influence learning as students tend to study what they know will be assessed. A summative assessment process can also have a formative component if it provides feedback to students on their performance. This learning experience can be stimulating and appreciated by students.

CONCLUSIONS

Whether for formative or summative purposes, student assessment is an essential step in the learning process. Each medical school must identify the most appropriate methods for assessing students for the various learning objectives: knowledge, skills, and attitudes. For those methods to be useful and fair, they must be developed and continuously improved to fulfil criteria such as validity, reliability, acceptability and feasibility.

The analysis of the results of student's assessment should guide medical schools in optimising teaching and learning processes. For students, it should contribute to their self-learning, so that they can play an active role in the process of their own improvement.

REFERENCES

1. Norcini JJ, & McKinley DW. Assessment methods in medical education. Teaching and teacher education. 2007 april;23(3):239-50. http://dx.doi.org/10.1016/ j.tate.2006.12.021.

2. Norcini JJ. Current perspectives in assessment: the assessment of performance at work. Med Educ. 2005 Sep;39(9):880-9. doi: 10.1111/j.1365-2929.2005.02182.x. PMID: 16150027.
3. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med J. 1975 Feb; 1(5955): 447-51. Available from: https://doi.org/10.1136/bmj.1.5955.447.
4. Epstein RM. Assessment in medical education. N Engl J Med. 2007 Jan 25;356(4):387-96. doi: 10.1056/NEJMra054784. PMID: 17251535.
5. Miller GE. The assessment of clinical skills/competence/ performance. Acad Med. 1990 Sep;65(9 Suppl):S63-7. doi: 10.1097/00001888-199009000-00045. PMID: 2400509.
6. Harden RM, Crosby JR, Davis MH, Friedman M. AMEE Guide No. 14: Outcome-based education: Part 5-From competency to meta-competency: a model for the specification of learning outcomes. Med Teach. 1999;21(6):546-52. doi: 10.1080/01421599978951. PMID: 21281173
7. McAleer S. Choosing assessment instruments. In: Dent JA, Harden RM, editor (2nd edition). A Practical Guide for Medical Teachers. London, UK: Elsevier Churchill Livingston; 2005. p 302-10.
8. Schuwirth LWT, van der Vleuten CPM. Written assessments. In: Dent JA, Harden RM, editor (2nd edition). A Practical Guide for Medical Teachers. London, UK: Elsevier Churchill Livingston; 2005. p 311-22.
9. Constructing Written Test Questions for the Basic and Clinical Sciences. Copyright © 2016 National Board of Medical Examiners® (NBME®).
10. McCoubrie P, McKnight L. Single best answer MCQs: a new format for the FRCR part 2a exam. Clin Radiol. 2008 May;63(5):506-10. doi: 10.1016/j.crad.2007.08.021. Epub 2008 Jan 14. PMID: 18374712.
11. Walsh JL, Harris BH, Smith PE. Single best answer question-writing tips for clinicians. Postgrad Med J. 2017 Feb;93(1096):76-81. doi: 10.1136/postgradmedj-2015-133893. Epub 2016 Jul 1. PMID: 27371033.
12. McCoubrie P. Improving the fairness of multiple-choice questions: a literature review. Med Teach. 2004 Dec;26(8):709-12. doi: 10.1080/01421590400013495. PMID: 15763874.
13. Klass D. Reevaluation of clinical competency. Am J Phys Med Rehabil. 2000 Sep-Oct;79(5):481-6. doi: 10.1097/00002060-200009000-00018. PMID: 10994893.
14. Marks M, Humphrey-Murto S. Performance assessment. In: Dent JA, Harden RM, editor (2nd edition). A Practical Guide for Medical Teachers. London, UK: Elsevier Churchill Livingston; 2005. p 323-35.
15. Davis MH, Ponnamperuma GG. Work-based assessment. In: Dent JA, Harden RM, editor (2nd edition). A Practical Guide for Medical Teachers. London, UK: Elsevier Churchill Livingston; 2005. p 336-45.
16. Norman G. The long case versus objective structured clinical examinations. BMJ. 2002 Mar 30;324(7340):748-9. doi: 10.1136/ bmj.324.7340.748. PMID: 11923143; PMCID: PMC1122692.
17. Norcini JJ, Blank LL, Duffy FD, Fortna GS. The mini-CEX: a method for assessing clinical skills. Ann Intern Med. 2003 Mar 18;138(6):476-81. doi: 10.7326/0003-4819-138-6-200303180-00012. PMID: 12639081.
18. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). Med Educ. 1979 Jan;13(1):41-54. PMID: 763183.
19. Hodges B, McIlroy JH. Analytic global OSCE ratings are sensitive to level of training. Med Educ. 2003 Nov;37(11):1012-6. doi: 10.1046/j.1365-2923.2003.01674.x. PMID: 14629415.
20. Patrício MF, Julião M, Fareleira F, Carneiro AV. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? Med Teach. 2013 Jun;35(6):503-14. doi: 10.3109/0142159X.2013.774330.
21. Gormley G. Summative OSCEs in undergraduate medical education. Ulster Med J. 2011 Sep;80(3):127-32. PMID: 23526843; PMCID: PMC3605523.
22. Reznick RK, Blackmore D, Cohen R, Baumber J, Rothman A, Smee S, Chalmers A, Poldre P, Birtwhistle R, Walsh P, et al. An objective structured clinical examination for the licentiate of the Medical Council of Canada: from research to reality. Acad Med. 1993 Oct;68(10 Suppl):S4-6. doi: 10.1097/00001888-199310000-00028. PMID: 8216627.
23. Davis MH, Harden RM. Competency-based assessment: making it a reality. Med Teach. 2003 Nov;25(6):565-8. doi: 10.1080/0142159032000153842. PMID: 15369903.
24. Driessen EW, van Tartwijk J, Overeem K, Vermunt JD, van der Vleuten CP. Conditions for successful reflective use of portfolios in undergraduate medical education. Med Educ. 2005 Dec;39(12):1230-5. doi: 10.1111/j.1365-2929.2005.02337.x. PMID: 16313582.
25. Davis MH, Ponnamperuma GG. Portfolios, projects and dissertations. In: Dent JA, Harden RM, editor (2nd edition). A Practical Guide for Medical Teachers. London, UK: Elsevier Churchill Livingston; 2005. p 347-56.
26. Smith PC, & Kendall LM. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of applied psychology.1963. 47(2): 149-55. https://doi.org/10.1037/h0047060
27. Likert R. A technique for the measurement of attitudes. Archives of Psychology. 1932. vol 140:1-55.
28. Guilford JP. Psychometric methods. 2nd ed. New York: McGraw-Hill; 1974
29. Falchikov N, & Boud D. Student Self-Assessment in Higher Education: A Meta-Analysis. RER. 1989. 59(4): 395–430. https:/ /doi.org/10.2307/1170205
30. Van Der Vleuten CP. The assessment of professional competence: Developments, research and practical implications. Adv Health Sci Educ Theory Pract. 1996 Jan;1(1):41-67. doi: 10.1007/BF00596229. PMID: 24178994.