DOI: https://doi.org/10.53350/pjmhs02023171156

**ORIGINAL ARTICLE**

# Evaluation Techniques in Medical Education: An Opinion

ATTAULLAH KHAN NIAZI[1], MUHAMMAD MUNEEB[2], MUHAMMAD BILAL MOEEN-UD-DIN[3], SADIA ZAHEER[4], MUHAMMAD MOHSIN MAHMOOD[5], MOHAMMED ALI[6], SALIK AHMED CHEEMA[7], YASIR RANA[8], BENAZEER MAKRANI[9], SAMAN QADEER[10]

[1]*Associate Professor Cardiac Surgery, King Edward medical university / Mayo Hospital Lahore*
[2]*Associate Professor, Department of Cardiology, Shalamar Medical & Dental College, Lahore*
[3]*Assistant Professor Medical Education, Sialkot Medical College, Sialkot*
[4]*Shifa College of Medicine, Shifa Tameer-e-Millat University, Islamabad*
[5]*Senior Registrar Cardiac Surgery, Mayo Hospital, Lahore*
[6-10]*Post graduate Trainee, Department of Cardiac Surgery, King Edward Medical University / Mayo Hospital, Lahore*
*Correspondence to: Attaullah Khan Niazi, drattaniazi@yahoo.com*

## ABSTRACT

Evaluation techniques play a crucial role in medical education at the undergraduate level. Medical educators have been exploring various evaluation techniques to ensure the comprehensive assessment of undergraduate medical students. In this article, we will delve into a comprehensive review and analysis of the existing evaluation methods used in medical education at the undergraduate level. By understanding the strengths and limitations of these techniques, educators and administrators can make informed decisions about the most effective evaluation methods for their institutions. Moreover, this review will provide valuable insights for future research and development in the field of medical education evaluation. To conduct a comprehensive review and analysis of evaluation techniques in medical education at the undergraduate level, a thorough literature review was conducted. The review included a wide range of sources such as academic journals, textbooks, and reports from reputable medical education organizations. Key evaluation methods were identified and analyzed based on their usage, effectiveness, and applicability to the undergraduate medical curriculum. Additionally, specific criteria such as validity, reliability, feasibility, and acceptability were considered in evaluating the effectiveness of these techniques. Furthermore, a comparative analysis of the identified evaluation techniques was performed, considering factors such as validity, reliability, feasibility, and acceptability among educators and students. This allowed for a deeper understanding of the strengths and limitations of each method, providing valuable insights into their practical implementation and impact on student learning outcomes.

**Keywords:** Evaluation techniques, medical education, applications, multiple choice question.

## INTRODUCTION

Miller introduced a model for assessing clinical competency which included several tiers of evaluation. The Derry 1990 model [1] emphasized the assessment of cognition leading to the evaluation of practice, with an increasing emphasis on professional authenticity as the assessment tasks move closer to real-world practice situations. Cognition was initially assessed based on knowledge and its application across various levels such as comprehension, application, analysis, and evaluation in line with Bloom's taxonomy. The assessment then moved towards evaluating behavior in controlled settings and actual performance using different tools best suited for each level. Van Vleuten also proposed a conceptual approach to determine the usefulness of evaluation instruments by weighting multiple criteria including validity, reliability, educational impact, acceptability, and cost depending on their purpose.[2] Summative evaluations focused more on reliability while formative evaluations prioritized educational impact when used for diagnosis, feedback, or improvement purposes.

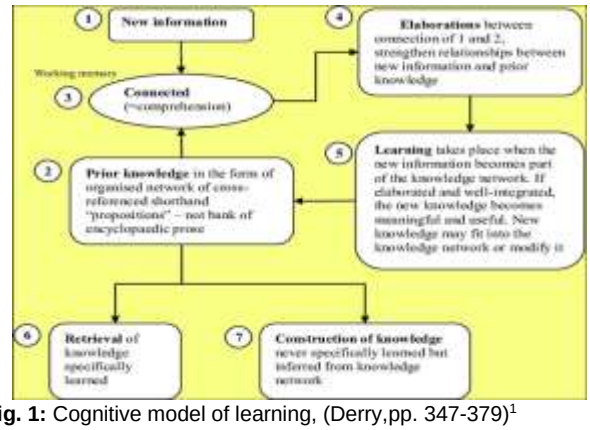


**Fig. 1:** Cognitive model of learning, (Derry,pp. 347-379)[1]

It is important to recognize that no single approach can cover all aspects of expertise in any field and therefore requires various assessment methods. The incorporation of multiple evaluation methods in the undergraduate medical curriculum ensures a comprehensive and well-rounded assessment of student learning and competency. Based on this extensive analysis and understanding of evaluation techniques, it is crucial to incorporate multiple assessment methods in the undergraduate medical curriculum to ensure a comprehensive and well planned assessment.

**Assessment of Knowledge and its Application:** Writing serves as a common method for assessing knowledge and can be conducted online, offering various written test formats. However, the quality of an exam is more reliant on the content of the questions rather than the format used for answering them. While some argue that multiple-choice questions may not effectively evaluate students' problem-solving abilities compared to open-ended questions, properly prepared MCQs can indeed assess these skills. It's essential to consider that certain question styles may be better suited for specific inquiries – such as using an essay question when the explanation is necessary. Evaluating both the advantages and disadvantages of each approach in a given set of questions is crucial since no single type could fully examine all attributes due to inherent limitations. To mitigate bias associated with individual formats, employing a variety of question types would be beneficial while ensuring alignment with program or course objectives.[3-4]

**1-Multiple choice questions:**

**a-(A-Type: One Best Answer)**

One of the advantages of using multiple-choice questions is their versatility in assessing different levels of learning outcomes, ranging from basic recall to higher-order thinking skills such as analysis and evaluation. Additionally, MCQs allow for efficient grading and provide a standardized means of assessment. However, it is important for educators to carefully construct the questions to ensure they align with the desired learning objectives and are free from ambiguity or potential misinterpretation. By utilizing well-designed multiple-choice questions, educators can effectively assess a wide range of content and objectives while providing an objective measurement of student ability.[5-6]

**b-Open-Ended Questions**

On the other hand, open-ended questions provide a more in-depth evaluation of students' understanding and critical thinking abilities. They allow for the expression of complex ideas and require students to demonstrate their knowledge coherently and articulately. While open-ended questions may be more time-consuming to grade, the insights gained from students' responses can offer valuable feedback on their comprehension and analytical skills.[7-8]

**c-Multiple Choice Question (R-Type: Extended matching questions)**

Extended Matching Questions are a valuable tool for contextual inquiries, containing sets of eight items organized into clinical scenarios with a single list of options focused on specific components. It is advised to include eight options instead of the usual five to six to minimize testing time. A well-constructed set includes four parts: lead-in phrase, options list, theme, and at least two item stems. Case and Swanson provide helpful tips for creating these queries. [9-10]

**2-Key feature questions**

The "key feature questions," are frequently utilized in clinical decision-making assessments. These queries revolve around brief clinical cases or scenarios and aim to evaluate the comprehension of essential case elements or critical judgments. They can take the form of open-ended or multiple-choice questions, with the potential for multiple correct answers from different individuals. Properly constructed key feature questions have been proven to possess validity and reliability based on specific guidelines. While used in high-stakes examinations in Canada and Australia, they are not as widely recognized as other question types. Nonetheless, creating them can be a time-consuming endeavor, especially for educators lacking expertise in question development.[11-12]

**3- Short Answer questions (SAQS)**

Short-answer questions require students to provide concise responses to specific prompts or questions that test their knowledge and understanding of the material. These questions can be effective in assessing students' ability to recall information and apply it to specific contexts. Short-answer questions require students to provide concise responses to specific prompts or questions, testing their ability to recall information and apply it to specific contexts. Careful crafting of the prompts is essential to ensure accurate evaluation of students' knowledge and skill.

**4- Essay Questions:**

Essay questions provide students with the opportunity to demonstrate their understanding of complex concepts and their ability to analyze, synthesize, and critically evaluate information. Essay-type questions are a valuable assessment tool in medical education, allowing for a more comprehensive evaluation of students' understanding, critical thinking abilities, and communication skills. These questions require students to provide detailed and coherent written responses, demonstrating their knowledge and ability to articulate complex ideas. In the context of medical education, essay questions can be particularly effective in assessing students' understanding of medical concepts, ethical reasoning, and decision-making skills. For example, students may be asked to analyze a clinical case study, discuss the implications of a medical scenario, or propose treatment plans based on their understanding of course content. Essay questions in medical education assessment provide valuable insights into students' clinical reasoning and communication skills. Clear prompts and effective grading criteria are essential for consistency and fairness. These questions assess students' understanding and critical thinking, preparing them for real-world practice.[11-13]

**5-MODIFIED ESSAY QUESTIONS (MEQS)**

Modified essay-type questions typically begin with a case followed by a series of related questions presented in chronological order. These questions are interdependent, meaning that an incorrect answer to one question may lead to subsequent inaccuracies. Students cannot revise previous responses and the case is reconstructed during grading, allowing no changes. MEQs primarily evaluate problem-solving, critical thinking, and comprehension skills rather than factual memory. Key feature questions are being used as replacements for multiple-choice questions to address psychometric issues associated with question interdependent questions, focusing on the assessment of higher-order thinking. The proper construction of MEQs is crucial to ensure that the case and related questions align with the learning objectives and desired outcomes. Additionally, thorough training and guidelines for graders are essential to ensure consistency and fairness in evaluating students' responses. MEQs offer a unique way to assess students' ability to apply knowledge in complex, real-world situations, making them a valuable addition to the assessment toolbox in medical education.[14]

**6- SCRIPT CONCORDANCE TEST (SCT)**

The Script Concordance Test is a valuable tool in assessing clinical reasoning skills in medical education. The SCT presents a clinical scenario followed by a question that requires the student to make decisions based on incomplete or uncertain information. The student must then rate the appropriateness of various actions or responses on a Likert scale.[15]

The Script Concordance Test is a relatively recent development in the field of health professions education and has been gaining popularity. This approach revolves around the concept that the various judgments made during clinical reasoning processes can be examined and evaluated based on their alignment with those of a panel of expert references. The SCT format is specifically designed to evaluate clinical reasoning within ambiguous situations, resembling the everyday challenges faced by physicians, lending it credibility. Brief Case Studies form part of SCTs by presenting short-case scenarios accompanied by related questions divided into three parts: relevant diagnostic or management options, new clinical findings, and a five-point Likert scale capturing examinees' decisions about how the new finding impacts the option.[15]

**7-The Objective Structured Clinical Examination (OSCE)**

In addition to these assessment methods, the Objective Structured Clinical Examination serves as a comprehensive tool for evaluating the basic clinical skills of students. The use of standardized patients, real patients, or simulators in different exercises helps to monitor and test specific competencies such as interpretation and technology skills.

The OSCE serves to evaluate the basic clinical skills of students through a series of "stations" that simulate different exercises. These stations can use standardized patients, real patients, or simulators to monitor and test specific competencies such as interpretation, non-patient competencies, and technology skills. The OSCE needs to have 14-18 stations to provide a credible performance metric. The scoring is done using work-specific checklists or rating scales, with global ratings providing results similar to checklists. Observers such as faculty, patients, or standardized patients can assign scores to trainees or students

In vitro performance evaluation involves assessments carried out under controlled conditions, while in vivo performance evaluation is based on real-world situations. It is important to note that these two types of evaluations are distinct from each other. To be evaluated in either category, a student must consistently demonstrate specific skills or conduct at particular intervals. The examiner observes and evaluates the student's abilities using various methods such as checklists, rating scales, structured reports, unstructured observations, as well as surveys like Objective Structured Clinical Examinations, and Direct Observation of Procedural Skills. However, the primary goal of clinical competency evaluation remains consistent: assessing physicians' performance in their medical practice widely regarded as the most dependable method for this purpose.[16-17]

**Checklist**

To develop a checklist, experts must reach a consensus on the primary behaviors, actions, and performance assessment criteria. This agreement is essential for accurate scoring and evaluation of the checklist. It's also vital to use checklists to ensure consistent ratings and to have evaluators who are properly trained in their usage since they serve as a reliable tool for evaluating personal and professional traits, competencies, and attitudes. Rating scales necessitate continuous or intermittent evaluations from observers which can be challenging and subjective. Given the limited reliability of rating scale decisions, it's important to ensure fairness by obtaining multiple independent ratings of the same student during the same activity. Moreover, providing training on rating applications is critical according to guidelines outlined by Davis and Ponnamperuma aimed at enhancing the quality of rating scales in assessments.

Feasibility is an important consideration when selecting an assessment format. The Script Concordance Test and Modified Essay Questions are valuable tools in assessing clinical reasoning and problem-solving skills in medical education. These methods provide a unique way to evaluate students' ability to apply knowledge in complex, real-world situations. Overall, a reliable and valid method for accurate assessment of clinical reasoning and problem-solving skills is crucial in medical education.

When developing assessment tools such as checklists and rating scales, it is important to ensure consistency and fairness in the evaluation process. Continuous training and guidelines for graders are essential to enhance the quality and reliability of the assessment tools. Furthermore, obtaining multiple independent ratings of the same student during the same activity can help ensure fairness and accuracy in the evaluation process.

Feasibility is an important consideration when selecting an assessment format, as it impacts the effectiveness and efficiency of the assessment process. It is crucial to carefully evaluate the feasibility of an assessment format to ensure that it aligns with the objectives and desired outcomes of the assessment while also being practical and manageable.

## CONCLUSION

The selection and development of assessment tools in medical education must consider the specific skills and competencies being evaluated, as well as the feasibility and reliability of the assessment format. Proper training and guidelines for graders are essential to ensure consistency and fairness in the evaluation process. Furthermore, the use of assessment methods such as the Script Concordance Test and Modified Essay Questions has been identified as beneficial in evaluating clinical reasoning and problem-solving skills in medical education. These methods offer a distinct approach to assessing students' abilities to apply knowledge in complex, real-world scenarios, contributing significantly to the overall assessment framework.

Additionally, the Objective Structured Clinical Examination continues to be a comprehensive tool for evaluating the basic clinical skills of students. Its use of various scenarios and standardized patients, real patients, or simulators provides a robust platform for testing specific competencies such as interpretation and technology skills.

When developing assessment tools like checklists and rating scales, it is paramount to prioritize consistency and fairness in the evaluation process. Continuous grader training and adherence to guidelines are essential in enhancing the quality and reliability of these assessment tools. Moreover, obtaining multiple independent ratings of the same student during the same activity can significantly contribute to the fairness and accuracy of the evaluation process.

Feasibility remains a critical consideration when selecting an assessment format, as it directly influences the effectiveness and efficiency of the assessment process. A thorough evaluation of the feasibility of an assessment format is crucial to ensure its alignment with the assessment objectives and desired outcomes while also being practical and manageable.

In conclusion, the careful selection and development of assessment tools in medical education should take into account the specific skills and competencies being evaluated, in addition to the feasibility and reliability of the assessment format. Adequate training and guidelines for graders are indispensable in maintaining consistency and fairness in the evaluation process, ultimately contributing to the overall success of medical education assessments.

## REFERENCES


1. Derry SJ. Learning strategies for acquiring useful knowledge. InDimensions of thinking and cognitive instruction 2013 May 13 (pp. 347-379). Routledge.
2. Schuwirth LW, Van der Vleuten CP. How to design a useful test: the principles of assessment. Understanding medical education: evidence, theory, and practice. 2018 Dec 3:275-89.
3. Kaipa RM. Multiple choice questions and essay questions in curriculum. Journal of Applied Research in Higher Education. 2021 Jan 20;13(1):16-32.
4. Haataja ES, Tolvanen A, Vilppu H, Kallio M, Peltonen J, Metsäpelto RL. Measuring higher-order cognitive skills with multiple choice questions–potentials and pitfalls of Finnish teacher education entrance. Teaching and Teacher Education. 2023 Feb 1;122:103943.
5. Ali K, Zahra D. Ten tips for effective use and quality assurance of multiple-choice questions in knowledge-based assessments. European Journal of Dental Education. 2024 Jan 28.
6. Mate K, Weidenhofer J. Considerations and strategies for effective online assessment with a focus on the biomedical sciences. Faseb Bioadvances. 2022 Jan;4(1):9.
7. Farhat H, Laughton J, Gangaram P, El Aifa K, Khenissi MC, Zaghouani O, Khadhraoui M, Gargouri I, Alinier G. Hazardous material and chemical, biological, radiological, and nuclear incident readiness among prehospital care professionals in the State of Qatar. Global Security: Health, Science and Policy. 2022 Dec 31;7(1):24-36.
8. Hauer KE, Boscardin C, Brenner JM, van Schaik SM, Papp KK. Twelve tips for assessing medical knowledge with open-ended questions: Designing constructed response examinations in medical education. Medical teacher. 2020 Aug 2;42(8):880-5.
9. Shin J, Guo Q, Gierl MJ. Multiple-choice item distractor development using topic modeling approaches. Frontiers in psychology. 2019 Apr 25;10:825.
10. Harris JL, Booth A, Cargo M, Hannes K, Harden A, Flemming K, Garside R, Pantoja T, Thomas J, Noyes J. Cochrane Qualitative and Implementation Methods Group guidance series—paper 2: methods for question formulation, searching, and protocol development for qualitative evidence synthesis. Journal of clinical epidemiology. 2018 May 1;97:39-48.
11. Hauer KE, Boscardin C, Brenner JM, van Schaik SM, Papp KK. Twelve tips for assessing medical knowledge with open-ended questions: Designing constructed response examinations in medical education. Medical teacher. 2020 Aug 2;42(8):880-5.
12. Pham H, Trigg M, Wu S, O'Connell A, Harry C, Barnard J, Devitt P. Choosing medical assessments: Does the multiple-choice question make the grade?. Education for Health. 2018 May 1;31(2):65-71.
13. Pushpalatha K, Kulkarni P, Manjunatha B. Assessment in competency based medical education: What, when and how?. Journal of Research in Medical Education & Ethics. 2020;10(2):60-9.
14. Gupta MM, Jankie S, Pancholi SS, Talukdar D, Sahu PK, Sa B. Asynchronous environment assessment: A pertinent option for medical and allied health profession education during the COVID-19 pandemic. Education Sciences. 2020 Nov 26;10(12):352.
15. Bhardwaj P, Black EW, Fantone III JC, Lopez M, Kelly M. Script Concordance Tests for Formative Clinical Reasoning and Problem-Solving Assessment in General Pediatrics. MedEdPORTAL. 2022 Sep 20;18:11274.
16. Vincent SC, Arulappan J, Amirtharaj A, Matua GA, Al Hashmi I. Objective structured clinical examination vs traditional clinical examination to evaluate students' clinical competence: A systematic review of nursing faculty and students' perceptions and experiences. Nurse Education Today. 2022 Jan 1;108:105170.
17. Loerwald AC, Lahner FM, Nouns ZM, Berendonk C, Norcini J, Greif R, Huwendiek S. 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 Jun 4;13(6):e0198009.


**This article may be cited as:** Niazi AK, Muneeb M., Moeen-ud-Din MB, Zaheer S, Mahmood MM, Ali M, Cheema SA, Rana Y, Makrani B, Qadeer S., Evaluation Techniques in Medical Education: An Opinion. Pak J Med Health Sci, 2023;18(11):56-58.