

Changing Assessment Scenarios: Lessons for Changing Practice

Abstract

Assessment is a process that includes ascertainment of improvement in the performance of students over time, motivation of students to study, evaluation of teaching methods, and ranking of student capabilities. It is an important component of the educational process influencing student learning. Although we have embarked on a new curricular model, assessment has remained largely ignored despite being the hallmark of competency-based education. During the earlier stages, the assessment was considered akin to “measurement,” believing that competence is “generic, fixed and transferable across content,” could be measured quantitatively and can be expressed as a single score. The objective assessment was the norm and subjective tools were considered unreliable and biased. It was soon realized that “competence is specific and nontransferable,” mandating the use of multiple assessment tools across multiple content areas using multiple assessors. A paradigm change through “programmatically assessment” only occurred with the understanding that competence is “dynamic, incremental and contextual.” Here, information about the students’ competence and progress is gathered continually over time, analysed and supplemented with purposefully collected additional information when needed, using carefully selected combination of tools and assessor expertise, leading to an authentic, observation-driven, institutional assessment system. In the conduct of any performance assessment, the assessor remains an important part of the process, therefore making assessor training indispensable. In this paper, we look at the changing paradigms of our understanding of clinical competence, corresponding global changes in assessment and then try to make out a case for adopting the prevailing trends in the assessment of clinical competence.

Keywords: *Assessment, assessor, competency-based medical education, faculty development program, measurement, programmatic assessment*

Introduction

Assessment is an important component of the educational process which influences students’ learning. Assessment is “the tail that wags the curriculum dog” meaning, “what is assessed is what gets learnt, which defines curriculum.”^[1] It is commonly believed that whatever is not “assessed” is not “possessed.” The very same dictum lures teachers to assess the students to such an extent that “learning” as a fundamental functional attribute of educational courses often gets replaced by “assessment” even in the absence of a robust and “fit for purpose” assessment system.

The word assess derives its name from the Latin word “*assidere*” meaning “to sit beside” and literally, “to assess” means “to sit beside the learner.”^[2] This actually means that assessment is something that we do “with” the students and not “to”

the students. “Assessment” in its simplest form uses various methods to document the learning of the students and is often inadvertently used synonymously with evaluation, judgment, measurement, appraisal, valuing, and rating, though there may be some context involved in using these terms. The assessment has been defined [Box 1] in various ways based on its role and purpose; ranging from “assessment to prove” to “assessment to improve.”

Currently, India is passing through an era of educational reforms-the National Education Policy has been notified recently;^[7] in the medical education arena, competency-based medical education has been introduced as a curricular model, aimed at producing the Indian Medical Graduate (IMG) who can address the community/societal health needs.^[8] These reforms in medical education mainly focus on teaching and learning methods, but assessment– in our opinion-seems to have

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Saiyad S, Bhagat P, Virk A, Mahajan R, Singh T. Changing assessment scenarios: Lessons for changing practice. *Int J App Basic Med Res* 2021;11:206-13.

**Shaista Saiyad,
Purvi Bhagat¹,
Amrit Virk²,
Rajiv Mahajan³,
Tejinder Singh⁴**

Department of Physiology, Smt N H L Municipal Medical College, 'M and J Western Regional Institute of Ophthalmology, B. J. Medical College, Ahmedabad, Gujarat, 'Department of Community Medicine, Adesh Medical College and Hospital, Kurukshetra, Haryana, 'Department of Pharmacology, Adesh Institute of Medical Sciences and Research, Bathinda, 'Department of Medical Education, Sri Guru Ram Das Institute of Medical Sciences and Research, Amritsar, Punjab, India

Submitted: 23-May-2021

Revised: 03-Aug-2021

Accepted: 02-Sep-2021

Published: 17-Nov-2021

Address for correspondence:

*Dr. Rajiv Mahajan,
Department of Pharmacology,
Adesh Institute of Medical
Sciences and Research,
Bathinda, Punjab, India.
E-mail: drrajivmahajan01
@gmail.com*

Access this article online

Website:
www.ijabmr.org

DOI:
10.4103/ijabmr.IJABMR_334_21

Quick Response Code:



been retained from the earlier curricular versions, with only cosmetic changes. Experts are unanimous in their opinion that competency-based curricula cannot be assessed by traditional assessment methods and assessment is one factor which separates traditional from competency-based curricula.^[9] Assessment in competency-based curricula should help in the development of competence rather merely detecting dyscompetence. It is therefore an ideal time to look-back at the evolution of different assessment concepts, methods and their transformation into their current form over time and global trends in their use, so that we evolve a “fit for purpose” robust and sustainable system of assessment in medical education in our country, which will support and augment the philosophy of competency-based curricula.

Changing Perspective of Clinical Competence

There has been a paradigm shift in our approach to assessment over time—from “measurement” to “decision making” to incorporation of a “program of assessment.” This transformation did not happen in vacuum as an isolated phenomenon. It was propelled by a concurrent change in our understanding of clinical competence [Figure 1]. Let’s look at the salient features of this change.

Measurement Era: The 1970s

Thorndike had asserted that “*anything that exists at all, exists in some quantity; and anything that exists in some quantity is capable of being measured.*”^[10] In the educational system, measurement refers to the quantification of learner performance in a given test and is often comparative. It is the process of assigning numbers to students or their characteristics without taking into consideration the value judgment that can be made from those numbers.^[11]

During the measurement era, the assessment was considered akin to assigning numerical values to the trait being assessed, and it was believed that competence can be measured purely quantitatively and even be expressed as a single score. During this era, “objective” assessment was promoted, and “subjective” assessment tools were viewed as being unreliable and biased. This led to the development of structured, standardized, reproducible, and objective methods for assessment. Attempts were made to minimize the role of human judgment by standardization and structuring to increase the reproducibility (often confused with reliability; discussed later) of an assessment.^[12]

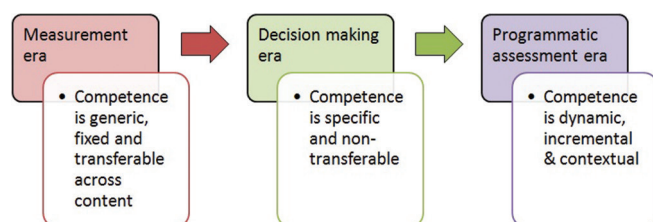


Figure 1: Changing perspectives about clinical competence

Reliability was regarded to be hallmarks of measurement and was considered as the inherent property of a tool.

This approach resulted in “objectification”—meaning an attempt to measure abstract concepts like competence as physical entities and thus designing a set of strategies to reduce measurement error. However, those *objectified* methods did not necessarily provide more reliable scores. On the contrary, it was recognized that objectified methods may induce unwanted outcomes such as triviality of content being measured and negative effects on learning behavior.^[13]

Several measurement models like Item Response Theory were developed to make the measurement more authentic.^[14] Most of these models presumed that the human mind works on a programmed mode and all students would perform as per the mathematical model. The reliability was however for consistency of *marking* and not for consistency of *performance*. Largely validity was sacrificed for reliability by atomizing and trivializing assessment.

It was also thought that the validity of a measure does not cover the validity of its use as operationalization of other target concepts.^[15] Thus, those standardized checklist-based psychometric tests were not sufficient to address all validity issues, and as such they did not measure what they were purported to measure.

Another dilemma which surfaced was that standardized psychometric measurements were based on limited samples of behavior and were subject to errors.^[16] Although single numeric scores gave precise measurement, they could not assess broad and complex competencies, especially those involving soft skills like communication, ethics, attitude, as well as those involving critical thinking. What could not be objectively measured was discarded from the assessment.

The measurement theorists considered competence as a generic trait, which would allow interpolation of results to other unrelated content. Using psychometric models for assessment of medical competence may not always be appropriate due to the element of case specificity i. e., unavoidable instability of performances across clinical cases or problems. Psychometric models assume *stability*

Box 1: Various definitions of assessment highlighting the role it plays in teaching and learning

Assessment refers to the processes employed to make judgments about the achievements of students over a course of study^[3]

The process of measuring an individual’s progress and accomplishments against defined standards and criteria, which often includes an attempt at measurement^[4]

Assessment is any formal or purported action to obtain information about the competence and performance of a student^[5]

Assessment in medical education addresses complex competencies and thus requires quantitative and qualitative information from different sources as well as professional judgment^[6]

of performance as central to assessment, which is not always the case, and which runs contrary to the premise of developmental nature of competencies. Another problem of standardized assessments was that to make tests equivalent for all students; the diagnostic, contextual, and inter-personal variables, which are part of authentic variability of real workplace settings, were discarded.^[17] For this reason, content and context specificity limited their applicability.^[18]

The opportunities of formative assessment and feedback were less in quantitative assessment as the main aim was summative analysis. This created a *reward-punishment* type of assessment which lacked crucial components of feedback. Assessment drives and affects learning. This is what Messick called *consequential validity*.^[19] Measurement type of assessment promoted test taking behaviors and superficial learning.

Despite having many drawbacks, measurement is still a good concept for assessment of the lower levels of Miller pyramid– at “knows” level. Multiple-choice questions, short answer questions, structured long answer questions, etc., depend heavily on measurement. In practice, quantitative measurements provide the idea about the overall achievement of the students but give no idea about the factors affecting the performance. In one perspective, this resembled a cross-sectional study, which did not allow the teachers and students to learn contextually. However, for norm-referenced testing and selection tests, measurement retains its value– not because of its inherent superiority but because of our programming that competence can be graded non-contextually, and a student scoring 63 is more competent than the one scoring 62!

It is interesting to note that none of the assessments are “purely” objective. Even the prototype of objectivity, the MCQs, also go through various stages like blueprinting, question design, choice of distractors and standard-setting– which are largely influenced by human interface and are subjective. Reducing cut-off scores, in case enough number of students have not qualified in selection tests is again a purely subjective exercise. Same is true of Objective Structured Clinical Examination. Various models have been developed for standard-setting, but again someone must tell the computer, what is an acceptable level. We have already elaborated on this issue.^[20]

Decision-Making Era: The 1990s

Limitations of using “measurement” as assessment prompted academicians to bring in the role of human judgment for assessing professional competencies, postpsychometric phase of assessment. Should we be assessing clinical competence like a race or like a gymnastic event? Though the concept of human judgment was subjective, it was necessary and much needed to assess important domains of professional competence such

as critical thinking, clinical reasoning, time management, teamwork, and doctor–patient relationship, many of which are an integral part of the conceptualization of an “Indian Medical Graduate.” Expert subjective judgment has shown its utility in many other domains. Before moving ahead, let’s consolidate our understanding about the concept of “assessment as measurement” and the felt need to move on to the concept of “assessment as decision making” through the example of the United States Medical Licensing Examination (USMLE).

USMLE serves many purposes like certification and selection for residency. For many years (till 2006), students’ grades were expressed as single numeric scores in the USMLE Step 1. This was a prime example of assessments in terms of measurement. It was presumed that knowledge can predict clinical performance (not a totally wrong presumption, influenced by content specificity, though); hence Step 1, which is related to assessment of the application of foundational science and is purely knowledge-based, continued in that form for many years, primarily guided by the need for objectivity and fairness. Many medical students prepared with “binge and purge” mentality, leading to short term retention only.^[21] Subsequently, the predictive value of numerically scored and only knowledge-based USMLE was found questionable for resident performance, except in the case of test failure.^[22] Though USMLE exams were framed within Messick’s conceptualization of construct validity, they could not meet validity criteria for secondary uses.^[23] It was therefore recommended to use other methodologies like assessment of clinical skills, for selection of candidates for residency programs leading to introduction of expertise driven, subjective, decision-making “Clinical Skills Assessment” in USMLE exam pattern, in addition to considering the students’ past performance.

Around the same time, two more concepts related to validity and reliability were propagated, eventually tilting the weight in favor of decision-making assessment. Messick gave the concept of *construct validity* as a unified and multi-faceted concept. According to this concept, all forms of validity are dependent and related to the quality of construct. The essence of unified validity is that the usefulness, meaningfulness, and appropriateness of score-based inferences are inseparable and the trustworthiness of empirically grounded score interpretation is the unifying force behind this integration.^[24] This was missing in pure quantitative assessments. Another important conceptualization was the observation that reliability is much more than reproducibility and that it does not co-vary with the objectivity of tools, thereby meaning that it is possible for objective tests to be unreliable and subjective tests to be reliable.^[13] Wide sampling across content and examiners were reported as the strategies to improve reliability of scores. With competence being defined as “the habitual and

judicious use of knowledge, communication, skills, clinical reasoning, emotional values and reflection in daily practice for the benefit of the individual and the society,”^[25] the case in favor of *longitudinal expert subjective* judgment was further built-up.^[20] This distinction can easily be understood by a simple example—seeing competence as a single trait was probably related to the folklore of using only a few grains of rice to decide if the broth was cooked. However, clinical competence now seems more like a complex biryani, where multiple ingredients need to be individually cooked and tested.

The assessment of competence and concept of quantification was certainly at loggerheads. *Shared subjectivity* was sometimes used to overcome this problem (e.g. four examiners assessing a single case and then coming to a consensus or using average regarding marks). However, this single shared perspective was not only not objective in the true sense, it also did not represent authentic practice, which is a range of perspectives on competence from all stakeholders.^[26] As such the concept of “shared subjectivity” no longer could stand the test of time, ultimately paving the way for the acceptance of the subjective.

Authentic assessment means observation and assessment of students while they are actually performing their work. This gives the assessor an opportunity to assess students in varying degrees of complexity in realistic situations. Though objectivity and standardization of an examination can increase reproducibility, there is risk of deviating from reality and authenticity, thus threatening validity though some, but not too much, the structure seems to improve the assessment process. Similarly, assessment literacy and expertise of the assessor and “adequate and representative” sample of tasks, tests, tools, and assessors were found to reduce the issues of subjectivity. The assessment of performance is a judgment and decision-making process; rating outcomes are affected by interactions between individuals and social context.^[27]

Decision-making era saw the growth of tools which relied on longitudinal expert subjective judgment—Mini Clinical Evaluation Exercise (m-CEX) and Professionalism mini evaluation exercise to name a few. It also saw the importance of assessing the student at the actual “place of work,” leading to the introduction of work-place based assessment (WPBA). The focus shifted to generalizability rather than reproducibility (i.e. how competent a student certified in one scenario will be in another and *not* if he will score the same marks again in the same scenario on a same case). Variability in assessment was accepted rather than frowned upon. Bigger sample size (assessors, assessments, content, and contexts) with some structure (likem-CEX) was the key element to improve generalizability and counter variability. WPBA was derived from the same measurement framework, and they became assessment processes that included human judgment based on assessment expertise.^[12]

With the introduction of CBME, the importance of expert subjectivity in students’ assessment has increased.^[20] Hallmarks of assessment in competency-based education are direct observation of trainees, feedback during formative assessment, and involvement of multiple assessors in multiple contexts. Though most of these assessment tools are subjective by common standards, validity and reliability can be increased by use of multiple encounters, multiple assessors and in multiple settings.^[28] The validity and reliability also shifted from the tool itself to *the way a tool is used*. Subjectivity helps in offering rich contextual feedback to learners rather than passing context-free judgments. Other essential attributes of CBME curriculum like reflective practice and self-directed learning are also promoted by subjective assessments. They provide opportunities to assessors to make defensible, fair, justifiable, clear, and learner-centered decisions.^[20]

Subjectivity in assessment cannot be avoided and so it should be recognized, adopted and used keeping all the checks to maintain validity, meaningfulness, and legitimacy of judgments. The bigger challenge is to build the rigor of quantitative data into subjective assessments.

Programmatic Assessment Era: The Mid-2000s

The mid-2000s witnessed another elementary change in the outlook towards assessment. This change was pursuant to a realization that education, competence, and assessment are complex entities, and an all-inclusive, comprehensive and holistic assessment is possible only at the level of a wholesome assessment program called programmatic assessment (PA).

This realization paved the way for a radical change in thinking around assessment with the following implications:

- Assessment in medical education encompasses observation, clinical reasoning, critical thinking, and decision-making activities, and all these may have diverse yet equally acceptable solutions
- Sub-optimal solution pathways mandate situational awareness, multiple strategies and a willingness to change swiftly.

Hence, it became essential to retrace and review the assessment process to facilitate its switch from a methods-based process to a comprehensive-system based process.^[12]

PA refers to an approach in which information about the student’s competence and progress is collected continually over time, analysed and is supplemented with purposefully collected additional assessment information, as and when needed.^[29] The aim is to inform the learner as well as the faculty and facilitate high-stakes decision making at the end of a training period.

PA approach rests on two notable and distinctive principles—the principle of proportionality and the principle of meaningful triangulation.^[30] PA entails an assessment

continuum that spans across individual low-stake assessments to high-stake decisions.^[29] This implies that all types of formal and informal assessments and feedback are low-stake that provide progress information specific to learning in all competency domains. The high-stake decisions require convincing interpretation of the results of a variety of assessment methods and rely on expert judgments of students' progress.^[31] The *principle of proportionality* thus refers to the percentage of abundantly rich information on student progress that contributes to making an assessment decision. In addition, in comparison to the conventional “one-tool-one-competency” assessment approach, in PA, the information collected from different sources contributes to all competency domains, and this is referred to as the *principle of meaningful triangulation*.

PA approach shifts the focus from individual assessments to a gamut of assessment instruments that are part of a larger whole. The need for expert subjective judgment in the assessment program and its utility in drawing meaningful conclusions from various assessment instruments is also in line with the contemporary viewpoint on validity which is not seen as an expression of numbers but a series of defensible narratives containing numbers as well as experimental outcomes.^[32] Validity and reliability of the entire assessment program add more value to assessment than the validity or reliability of individual assessments both in terms of the variety of methods that can be used and also assessment of competencies that are currently overlooked.^[33] PA theoretically aligns well with the goals of competency-based medical education.^[34] Shift to the classical model of PA requires a lot of change involving all stakeholders, especially regarding the utility of subjective assessments and may not be feasible immediately: We have already elucidated on the possibility of implementation of a blended PA for competency-based curricula in India.^[35]

Changing Role of Assessor

Conventionally, assessments place a high level of confidence upon the abilities of a second party for judgment. In the conduct of any assessment, the assessor is as much an integral part of the assessment as the learner.^[36] Even though most teachers are well-versed with the summative or “assessment of learning,” using assessment as an educational tool (assessment for learning) is a relatively recent phenomenon. The “*one instrument-one competency*” approach has now been replaced by a “*multiple-instruments-multiple competencies*” approach. Students' assessment is a program in itself rather than an addendum to the teaching-learning program. The role of assessor is now on the center stage, whose expertise decides the quality of assessment.

Schuwirth and Vleuten^[37] have illustrated it well by exemplifying that in the measurement phase, “assessors focused on demonstrating that a “hammer is better than a screwdriver” for certain tasks. In the competencies phase,

the purposes, pros and cons of hammers and screwdrivers were studied for their utility. Finally, in the PA era, value shifted to the combination of quality of the hammer and screwdriver (affordance of the tool) and expertise of the carpenter (effectivities of the user)”. This is aptly illustrated in the concept of the assessment toolbox, and we have already elaborated on this toolbox concept earlier.^[38] The evolution of assessment in medical education has been briefed in Table 1.

The Way Forward

It is pertinent to remember that the competence is contextual, constructed, and changeable, and hence is subjective and collective.^[17] Competency manifests in performance and does not reside in the individual; rather, it resides in the way an individual interacts with the context.

Assessments do not happen in a vacuum— they are contextual and for a purpose. While psychometric concepts are desirable for very high-stake assessments like selection and certification, we are denying the students the benefits of “assessment for learning” by insisting on the same standards even for class tests and formative assessments. One of the important changes in the new regulations is the incorporation of purely formative assessments, which provide us with a wonderful opportunity.^[39] We must exploit it to promote learning.

However, one should be aware that “subjective” assessments are based on or influenced by personal knowledge, expertise, beliefs and opinion of assessors. Experts too can make poor judgments. The teacher's experience therefore becomes the assessment “instrument.” Furthermore, assessment ability is acquired and not innate.^[40] Assessor training is hence an indispensable part of any assessment program. The areas where a teacher's capacity building needs to be addressed include making aware of one's limits as assessor and possibilities of being biased, identifying the needed competencies, their

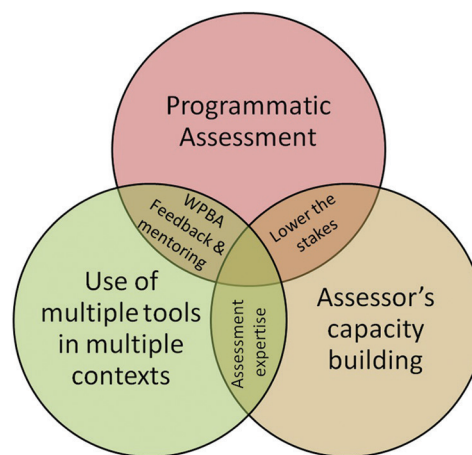


Figure 2: Futuristic approach for an authentic competency assessment system

Table 1: Evolution of assessment in medical education

Salient features	Key points	Limitations	Implications	Role of assessor
The measurement phase (1970s)				
Learner competence purely a quantitative measure; can be expressed as a (single) score Competence is generic and transferable Aimed at minimizing the role of human judgment to reduce unreliability Structuring and standardization to increase reliability	Objectivity important for reliability Stimulus format more important than response format Validity can be built in by quality assurance in item and test development Mathematical models can predict competence and behavior Assessment drives learning	Numerical results from structured, standardized tests insufficient to determine student competence Difficult to measure inter-dependent traits Student performance did not generalize well across content Objectivity did not ensure reliability/generalizability	Measurement still useful for lower levels of Miller and selection tests “Does” part not assessable by pure numbers	Objectification and standardization to compensate for “noise” in assessment Assessor training to design standardized tools Can produce high inter-rater agreement Validity and reliability ‘built into’ the tool
Decision making phase (1990s)				
Competence is not transferable Assessment of performance is a judgment and decision-making process Asserted the role of human judgment in the assessment process Assessment in real authentic settings inclusive of aspects such as critical thinking, professionalism, reflection and self-regulation	Bias is inherent to expert subjective judgment Manipulating numbers do not manipulate reality Validity and reliability not inherent to the tool Qualitative information has more value than mere scores. More so, for complex domain-independent skills Feedback is an indispensable component of assessment Assessment drives learning	Inter-rater agreement may be low Resource and time intensive May not be appropriate for selection type tests Learner has to be part of assessment process Systems and structures may be unsupportive Stakeholders buy in needed	Expert judgment is indispensable to assessing complex skills Devise strategies to alleviate bias Create meaningful relationship between teacher and learner through feedback dialogues and follow-up Capacity building of assessors in use of assessment tools	Use appropriate and adequate sampling of assessors to dilute bias Feedback based on direct observation Assessor training important to pick up dyscompetence Validity and reliability depend on the way a tool is used
Programmatic assessment phase (mid - 2000s)				
Signifies an approach in which information about the learner’s competence and progress is collected continually over time, analyzed and, as and when needed, supplemented with purposefully collected additional assessment information	No single method can assess all levels of Miller’s pyramid Whole task assessment Delinking of assessment and decision making Concept of utility of assessment with meaningful compromises and aggregation Learning drives assessment	Need of expert judgment to make meaningful conclusions from the information gathered using a different assessment instrument Need of examiners with sufficient assessment literacy and expertise Systems and structures may be unsupportive Stakeholders buy-in needed	Optimizes decision making by gathering rich and abundant information from multiple assessment times, methods and sources Assessment used purposefully to achieve the desired learning impact on learners Different examiners with different thought processes add richness to assessment	Multiple assessors in multiple settings/ contexts Assessor must be domain expert Feedback and mentoring Validity and reliability of the “program” rather than of the tool

performance expectations, measurement criteria in all relevant domains, design, and use of assessment tools and giving feedback to learners. The main hurdles for this are the basic lack of awareness about the need and required “knowhow” of assessment, the need for change in teacher-learner ethos and availability of resources.^[41] Assessor training must always be feasible and meaningful,

and integrated into the ongoing faculty development program.^[40] In fact, in the program of assessment, the use of carefully selected combination of tools and assessor expertise developed through faculty capacity building must be integrated for an authentic, observational driven, workplace-based institutional assessment system to be in place [Figure 2].

Healthcare providers are now becoming increasingly dependent on technology thereby necessitating future professionals to have newer yet different skills, abilities, and competencies. This, however, would require a new rethink of assessment and continued assessor training.^[12] The importance of content, contexts, tasks, settings, and assessors can never be overemphasized if we have to design competency-based assessments. Since competence is contextual, the assessments must be broad-based enough to make them generalizable. It is good to remember that we certify students as fit to practice medicine without any fine print saying, “terms and conditions apply.” Specifying competencies but not assessing them would be counterproductive or even dangerous. Each competency needs to be assessed at some point of time-formative, internal or summative; we cannot afford to define a competency and not assess it.

Conclusion

Development in the assessment arena has progressed from measurement to a human judgment perspective.^[37] Just as we need multiple assessment methods, we need multiple assessors to compensate for shortcomings such as biases, halo effects, and leniency so that the true picture is visible.^[40] For authentic assessment to happen, a program of assessment must be in place which must be aligned with the assessor’s expertise to use a tool from the assessment toolbox. The assessment shall see a paradigm shift, in the nature of tools used, in the way it shall be used, and inferences which will be drawn from it.^[38]

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

- Sergiovanni T, Starratt R. Supervision: A Redefinition. Boston: McGraw Hill; 2007. p. 127.
- Stefanakis EH. Multiple Intelligences and Portfolios: A Window into the Learner’s Mind. Portsmouth, NH: Heinemann; 2002. p. 9.
- Harlen W. Teachers’ summative practices and assessment for learning- tensions and synergies. *Curric J* 2005;16:207-23.
- Postgraduate Medical Education and Training Board, London. Developing and Maintaining an Assessment System – A PMETB Guide to Good Practice. London: Postgraduate Medical Education and Training Board; 2007. p. 46.
- Schuwirth LW, van der Vleuten CP. How to design a useful test: The principles of assessment. In: Swanwick T, Forrest K, O’Brien BC editors. *Understanding Medical Education: Evidence, Theory and Practice*. 3rd ed. Oxford, UK: Wiley Blackwell; 2019. p. 277.
- van der Vleuten CP, Schuwirth LW. Assessing professional competence: From methods to programmes. *Med Educ* 2005;39:309-17.
- Government of India: Ministry of Human Resource Development. National Education Policy 2020. Available from: https://niepiid.nic.in/nep_2020.pdf. [Last accessed on 2021 May 12].
- National Medical Commission. Competency Based Undergraduate Curriculum. Available from: <https://www.nmc.org.in/information-desk/for-colleges/ug-curriculum>. [Last accessed on 2021 May 12].
- Schuwirth L, Ash J. Assessing tomorrow’s learners: In competency-based education only a radically different holistic method of assessment will work. Six things we could forget. *Med Teach* 2013;35:555-9.
- Thorndike EL. The nature, purposes, and general methods of measurement of educational products. In: *The Seventeenth Yearbook of the National Society for the Study of Education*. Bloomington IL: Public School Publishing Company; 1918.
- Ebel RL. *Essentials of Educational Measurement*. 3rd ed. Englewood Cliff: Prentice-Hall; 1979.
- Schuwirth LW, van der Vleuten CP. A history of assessment in medical education. *Adv Health Sci Educ Theory Pract* 2020;25:1045-56.
- van der Vleuten CP, Norman GR, De Graaff E. Pitfalls in the pursuit of objectivity: Issues of reliability. *Med Educ* 1991;25:110-18.
- Lord FM. *Applications of Item Response Theory to Practical Testing Problems*. 1st ed. New York: Routledge; 1980.
- Truijens FL, Cornelis S, Desmet M, De Smet MM, Meganck R. Validity beyond measurement: Why psychometric validity is insufficient for valid psychotherapy research. *FrontPsychol* 2019;10:532. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6423000/>. [Last accessed on 14 May 2021].
- Crocker L, Algina J. *Introduction to classical and modern test theory*. New York: CBS College Publishing; 1986.
- Hodges B. Assessment in the post-psychometric era: Learning to love the subjective and collective. *Med Teach* 2013;35:564-8.
- Schauber SK, Hecht M, Nouns ZM. Why assessment in medical education needs a solid foundation in modern test theory. *Adv Health Sci Educ Theory Pract* 2018;23:217-32.
- Messick S. The interplay of evidence and consequences in the validation of performance assessments. *Educ Res* 1994;23:13-23.
- Virk A, Joshi A, Mahajan R, Singh T. The power of subjectivity in competency-based assessment. *J Postgrad Med* 2020;66:200-5.
- Haist SA, Katsufraakis PJ, Dillon GF. The evolution of the United States Medical Licensing Examination (USMLE): Enhancing assessment of practice-related competencies. *JAMA* 2013;310:2245-6.
- Hartman ND, Lefebvre CW, Manthey DE. A narrative review of the evidence supporting factors used by residency program directors to select applicants for interviews. *J Grad Med Educ* 2019;11:268-73.
- George P, Santen S, Hammoud M, Skochelak S. Stepping back: Re-evaluating the use of the numeric score in USMLE examinations. *Med Sci Educ* 2020;30:565-7.
- Messick S. *Foundations of Validity: Meaning and Consequences in Psychological Assessment*. Princeton, NJ: Educational Testing Service; 1993. Available from: <https://onlinelibrary.wiley.com/doi/epdf/10.1002/j.2333-8504.1993.tb01562.x>. [Last on accessed 14 May 2021].
- Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA* 2002;287:226-35.
- Ten Cate O, Regehr G. The power of subjectivity in the assessment of medical trainees. *Acad Med* 2019;94:333-7.
- Govaerts MJ, van der Vleuten CP, Schuwirth LW, Muijtjens AM. Broadening perspectives on clinical performance assessment: Rethinking the nature of in-training assessment. *Adv Health Sci Educ Theory Pract* 2007;12:239-60.
- Singh T, Sood R. Workplace-based assessment: Measuring and shaping clinical learning. *Natl Med J India* 2013;26:42-6.
- van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J,

- Tigelaar D, Baartman LK, *et al.* A model for programmatic assessment fit for purpose. *Med Teach* 2012;34:205-14.
30. Schut S, Maggio LA, Heeneman S, Tartwijk JV, van der Vleuten CP, Driessen E. Where the rubber meets the road – An integrative review of programmatic assessment in healthcare professions education. *Perspect Med Educ* 2021;10:6-13.
 31. Schuwirth LW, Van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. *Med Teach* 2011;33:478-85.
 32. Schuwirth LW, van der Vleuten CP. Programmatic assessment and Kane's validity perspective. *Med Educ* 2012;46:38-48.
 33. Singh T. Student assessment: Moving over to programmatic assessment. *Int J Appl Basic Med Res* 2016;6:149-50.
 34. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach* 2010;32:676-82.
 35. Mahajan R, Saiyad S, Virk A, Joshi A, Singh T. Blended programmatic assessment for competency based curricula. *J Postgrad Med* 2021;67:18-23.
 36. Gallagher P. The role of the assessor in the assessment of practice: An alternative view. *Med Teach* 2010;32:e413-6.
 37. Schuwirth LW, van der Vleuten CP. How 'Testing' has become 'Programmatic Assessment for Learning'. *Health Prof Educ* 2019;5:177-84.
 38. Singh T, Saiyad S, Virk A, Kalra J, Mahajan R. Assessment toolbox for Indian medical graduate competencies. *J Postgrad Med* 2021;67:80-90.
 39. Medical Council of India. Regulations on Graduate Medical Education (Amendment); 2019. Available from: <https://www.nmc.org.in/ActivitiWebClient/open/getDocument?path=/Documents/Public/Portal/Gazette/GME-06.11.2019.pdf>. [Last accessed on 2021 May 12].
 40. Lockyer J, Carraccio C, Chan MK, Hart D, Smee S, Touchie C, *et al.* Core principles of assessment in competency-based medical education. *Med Teach* 2017;39:609-16.
 41. Chacko TV. Moving toward competency-based education: Challenges and the way forward. *Arch Med Health Sci* 2014;2:247-53.