CHAPTER 9

A model for selecting assessment methods for evaluating medical students in African medical schools.[11]

Abstract

Introduction. The introduction of more effective and standardized assessment methods for testing student performance in African medical schools has been hampered by severe financial and personnel shortages. Nevertheless, some African institutions have recognized the problem and are revising their medical curricula, and, therefore, their assessment methods. These institutions, and those yet to come, need guidance on selecting assessment methods so as to adopt models that can be sustained locally.

Purpose. Develop an assessment utility model suitable for use in African medical schools.

Methods. The authors provide a model for selecting assessment methods for testing student performance in African medical schools. The model systematically evaluates four factors that influence the selection and implementation of assessment methods. Six commonly used methods – essay questions (EQ), short-answer questions (SAQ), multiple-choice questions (MCQ), patient-based clinical examinations (PCE), problem-based oral examinations (POE) and objective structured clinical examinations (OSCE) – were evaluated by calculating a score weighted for performance, cost, suitability, and safety factors. In the model, the highest score identified the most appropriate method.

Results. Selection of an assessment method is illustrated using two institutional models depicting (a) an ideal situation in which the OSCE was identified as the preferred method and (b) the typical African scenario in which EQ or SAQ methods were preferred. Improved resource allocation in Africa could make the POE method the most appropriate for African

[11] Walubo A, Burch V, Parmar P, Raidoo D, Cassimjee M, Onia R, Ofei F. A model for selecting assessment methods for evaluating medical students in African medical schools. Academic Medicine 2003; 78: 899-906.

medical schools, but this may require changing medial curricula to a problem-based learning approach.

Conclusion. The authors' model is easy to understand and suggests objective ways of improving the selection of student assessment methods in Africa. The model may also promote changes in medical school curricula in resource-constrained environments.

Introduction

Selecting an assessment method for measuring student performance remains a daunting task for many medical institutions in Africa where attempts to revamp existing assessment methods have been hampered by serious financial and personnel shortcomings and the lack of suitable alternatives.[1] The difficulty of this task is further compounded by the fact that changes in assessment methods may require changes in medical curricula.[2] Therefore, in Africa, measuring student performance is not the sole determinant of assessment method selection. Other factors such as cost, suitability, and safety have profound influences on the selection of appropriate assessment methods and are, most probably, the major cause of inter-institutional variation not only in the choice of assessment methods but also in student success rates.[3] Even at institutions where several assessment methods are used, their integration has been greatly influenced by the aforementioned factors. We feel strongly that the use of a comprehensive objective method of evaluating the performance, cost, suitability, and safety of assessment methods should be a prerequisite for appropriate assessment method selection.[4-6]

Unfortunately, assessment methods used by most medical institutions in Africa have been selected empirically and, hence, are highly variable and inconsistent both intra-institutionally and inter-institutionally. This situation may have arisen from difficulties in objectively evaluating potential influencing factors, and it highlights the need for a framework guiding the objective selection of assessment methods. Such a selection framework is especially needed in South Africa and other countries in Africa where medical schools are transforming their learning curricula and, therefore, their assessment methods. A lack of guidance on the selection of assessment methods may result in institutions adopting testing models that cannot be locally sustained. The paucity of literature regarding the selection and use of assessment methods appropriate to the African context underscores the serious nature of the problem. Providing a context for the selection of assessment methods that is specific for African medical institutions would give due regard to the extremely limited human resources in Africa, which are further confounded by adverse socioeconomic, fiscal and political factors. In this article, using our experience as both trainees and trainers in Africa, we propose a standard approach for selecting assessment methods for testing student performance in African medical schools. Although this proposal was stimulated by the need to guide African institutions in the selection of assessment methods appropriate to their resource constraints, we believe the selection criteria be useful to other developing world institutions, and even medical schools in the developed world.

Methods

Assessment methods evaluated

We compared six assessment methods: essay questions (EQ), short-answer questions (SAQ), multiple-choice questions (MCQ), patient-based clinical examinations (PCE), problem-based oral examinations (POE) and objective structured clinical examination (OSCE) for their ability to test student performance and their ease of adoption with regard to cost, suitability and safety. Each of these factors is described below, as are the rating scales we used to evaluate them. We used our evaluation on a standardized class size of 150 students. The six assessment methods selected were used to evaluate the utility of this model because they are widely used in most African medical institutions.

Description of assessment methods

To avoid complexities that might arise from minor inter-institutional variations in administering the assessment methods, we standardized the description of each method and examination time for our evaluation process. Essay questions traditionally require lengthy written responses to one or more questions,[7,8] while SAQ involve short written responses to a greater number of questions sampling more broadly from the curriculum.[9] Both EQ and SAQ examinations are usually two hours in duration and are scored out of 100 marks each. Preparing these examinations involves setting the questions and a memorandum of answers which must be reviewed by a moderator. MCQ questions usually require selection of an answer from a choice (best option) and preparation of an MCQ examination has similar requirements to the other written tests described. MCQ papers are usually marked by computer and do not require a memoranda. They are usually of similar duration, 1.5 to 2 hours.

The PCE is usually a two-hour examination involving one long case and two short cases using non-standardized patients. The long case tests students' skills in patient clerking (interview and patient examination) and takes 1.5 hours (45 minutes for clerking, 15 minutes for presentation, and 30 minutes for evaluation). The two short cases test students' skills in examination techniques and diagnosis, and each lasts for 15 minutes. All examination patients are selected at random, are not previously seen by the student, and may be different for different students.

The POE is a two-hour examination which primarily focuses on problem-solving skills and clinical reasoning. Students are given an hour to study and evaluate a paper-based case description and write appropriate notes regarding the case. After the long case has been marked, students are examined in a one-hour oral (viva) examination. The oral examination discusses the long case and two additional paper-based short case scenarios (30 minutes covers the long case

and 30 minutes covers two short cases). The short cases are picked at random, are not previously known to the student, and may be different for different students.

The OSCE is a two-hour examination during which students move through several stations (10 minutes per station x 10 stations) where they are examined on different aspects of the station's subject or clinical materials put out at the station, e.g. chest radiographs or electrocardiographic (ECG) recordings. The effective examination time is 100 minutes, with 20 minutes allowed for moving between stations.

The PCE, POE, and OSCE each require several examination stations or rooms with at least three examiners per room.

Model for evaluating assessment methods

Performance. We scored each assessment method for performance in terms of the method's ability to test clinical problem solving skills, knowledge and recall of facts, communication and practical (clinical examination) skills. The maximum performance score for each assessment method was 10 points. Because problem-solving skills are considered a fundamental outcome of medical training, we weighted it more (a possible four points) than we weighted the other three skills (a maximum of two points each). If an assessment method reliably tested clinical problem-solving skills we scored it as four; one with a limited scope for testing this skill was scored two. Methods with limited problem solving testing capacity scored one, and methods not testing problem-solving skills were scored zero. A method with the ability to test for knowledge (or to sample a wide range of knowledge) was scored two, whereas one with a limited ability to test knowledge (or a limited sampling range) was scored one; tests not assessing knowledge in any significant way scored zero. The evaluation of each method's ability to test communication and practical skills were scored along the same scale as the knowledge parameter already outlined.

Cost. We evaluated the cost of implementing an assessment method based on the cost of materials required for examination purposes, examiners (paid according to hours of examination and marking), and patients or patient simulators. Again, we used a 10-point scale for this dimension of the model. The cost of examiners was weighted the most (six points); other materials and patient costs were separately scored – a maximum of two points each.

The cost of an examiner was evaluated in terms of the number of manpower hours (MPH); that is, the time required to complete the whole examination process, including examination supervision and marking. MPH excluded preparation of examinations, which we considered a suitability factor (described later). The scale for scoring for MPH was as follows: 0-10 MPH = 6; 11-100 MPH = 5; 101-200 MPH = 4; 201-300 MPH = 3; 301-400 MPH = 2; 401-500 MPH = 1; and more than 500 MPH = 0.

We defined extra materials to include stationary (pens, paper, MCQ answer cards), textbooks, computers, videos, x-ray films, specimens, laboratory test results, etc. If an assessment method required only stationary, it was scored two. If it required one extra item in addition to stationary, it was scored one; if it required two or more extra items other than stationary, it scored zero.

We defined patient costs as expenditures, direct and indirect, on patients or patient simulators kept in hospital for examination purposes. The need for patient participation was scored separately so as to enable application of the selection model to both clinical and non-clinical examinations. Also, separating these scoring components emphasized the resource implications of using patients or patient simulators. More importantly, the need for patients distinguished the POE from the PCE or OSCE. We scored an assessment method that did not involve the use of patients or patient simulators as two. If the institution incurred only the indirect cost of extended patient stays for assessment purposes it was scored one; methods using patients or patient simulators were scored zero.

Suitability. We scored an assessment method's suitability to the institution based on the time required to prepare the examination, the need for an examination venue (cubicle hours), and the need for special examination requirements at the venue, termed resources. The maximum possible score was 10. We considered examination preparation time a suitability factor because, in addition to setting questions, it often involves identifying suitable patients, preparing the venue and other special examination arrangements.

The venue was considered the most important suitability factor (maximum score of six), and resources required and examination preparation each scored a maximum of two points. Venue requirements were calculated as total cubicle hours (CH) required for examining all 150 students: 0-10 CH = 6; 11-100 CH = 5; 101-200 CH = 4; 201-300 CH = 3; 301-400 CH = 2; 401-500 CH = 1; and more than 500 CH = 0.

Resources beyond those normally used in training, such as a table and chair for every student, were thought to include beds, shield curtains, major modification of rooms or environment, etc. If an assessment did not call for special resources, it was scored as two; an addition of one extra resource was scored one, and an addition of two or more resources was scored zero.

Examination preparation was expressed as effective work-hours. If the examination preparation took less than 12 effective work-hours, it was scored two, 12 to 24 hours was scored one and more than 24 hours was scored zero.

Safety. Safety or proof of an assessment method to resist leakage prior to, or during, the examination and cheating (student communication during the examination) of any nature were

scored. A score of six was given to an assessment method where communication between students was deemed impossible. If communication was possible but unlikely to improve student performance, the method received a score of four. When communication could indirectly influence student performance, it received a score of two. Examinations such as project execution tests and portfolios, where communication between students may form part of the learning and assessment process, were scored as zero. These types of assessment did not, however, form part of the evaluation described in this paper.

For test leakage, we scored assessment methods as follows: those that alluded to any probability of leakage were scored zero, those that had no probability of leakage, but examiner bias could significantly influence student performance were scored two, and those with no probability of leakage or examiner bias were given a score of four.

Results

Performance

Table 1 shows the scores for the four parameters of student performance. Overall, the best method for assessing student clinical performance was the OSCE; the MCQ faired the worst. The PCE and POE assessment methods, which have a limited capacity to test knowledge and recall of facts, together with the OSCE, received high scores for problem-solving, communication, and practical skills.

Table 1. *Performance parameters for six assessment methods used at African medical schools*

	Test performance (Maximum score)				
Assessment method	**Knowledge and recall of facts (2)**	**Problem solving (4)**	**Communication (2)**	**Physical examination (2)**	**Total score (10)**
EQ	2	2	1	1	6
SAQ	2	2	1	1	6
MCQ	2	1	0	0	3
PCE	1	4	2	2	9
POE	1	4	2	1	8
OSCE	2	4	2	2	10

Cost

Table 2 shows the cost parameters of the six assessment methods evaluated. Scores for cost of examiners, examination materials, and patients were evaluated. Overall, MCQ tests cost

the least due to computerized marking. The total examination time, 9.5 MPH, received a score of six. One extra examination resource (computer) was required, but patients were not applicable.

Table 2. *Cost parameters for six assessment methods used at African medical schools*

	Test cost (Maximum score)			
Assessment method	**Examiners (6)**	**Materials (2)**	**Patients (2)**	**Total score (10)**
EQ	3	2	2	7
SAQ	3	2	2	7
MCQ	6	1	2	9
PCE	0	0	1	1
POE	0	2	2	4
OSCE	0	0	0	0

The EQ and SAQ papers differed marginally in estimated marking time (1.5 hours per student versus 1.3 hours per student, respectively). Both required five supervisors (two hours each) and required no additional materials (students only needed standard stationary). Total examination time for the PCE was 675 MPH, giving a score of zero, and extra examination materials were required (i.e. x-rays and laboratory investigation results). For the purpose of the analysis, non-paid in-patients were used. Total examination time for the POE was 530 MPH, giving a score of zero. The cost of examination materials and patients were not applicable, as students only needed standard stationary. The OSCE was the most costly examination. With 10 stations and three examiners per session (three different stations), 150 students would require 900 MPH, achieving a score of zero. More than one extra examination resource was required (i.e. computers, photographs, videos recorders, x-rays, laboratory investigation results and specimens), and paid patients simulators were calculated as part of the cost.

Suitability

Table 3 shows the suitability scores of assessment methods evaluated. In general, EQ, SAQ, and MCQ examinations obtained the highest suitability scores. Test methods involving patients, i.e. PCE and OSCE, required a lot of venue time, examination preparation, and many additional resources. In the case of EQ, SAQ, and MCQ examinations, all 150 students could be accommodated in one venue at one time. They respectively required less than 10 CH, for a score of six each. Furthermore, they required no additional resources beyond chairs and tables. The PCE required two CH per student (300 CH in total) and more than one extra resource (i.e. shield curtains and beds). Examination preparation times exceeded 24 work-hours. The POE

required one CH per student (151CH, including supervision of one hour for long case write up), but no special resource beyond that of chairs and tables. Preparation time was less than 12 hours. The OSCE requires two CH per student (a total of 300 CH), and more than one extra resource (i.e. beds and shield curtains or other modification of rooms to make examination stations). Examination preparation time also exceeded 24 work-hours.

Table 3. *Suitability parameters for six assessment methods used at African medical schools*

	Test suitability (Maximum score)			
Assessment method	**Examiners (6)**	**Materials (2)**	**Patients (2)**	**Total Score (10)**
EQ	6	2	2	10
SAQ	6	2	2	10
MCQ	6	2	2	10
PCE	3	0	0	3
POE	4	2	2	8
OSCE	3	0	0	3

Safety

Table 4 contains examination safety scores for the assessment methods. Communication between students is generally not possible during EQ, SAQ, and MCQ examinations. Whereas it is true that some students may discuss their PCE, POE, and OSCE examination experiences, there is no proof that this communication improves student performance. The probability of examination leakage is higher for examinations where questions are predetermined, including EQ, SAQ, MCQ and POE.

Table 4. *Safety parameter scores for six assessment methods used at African medical schools*

	Test safety (Maximum score)		
Assessment Method	**Communication between candidates (6)**	**Probability of test leakage (4)**	**Total score (10)**
EQ	6	0	6
SAQ	6	0	6
MCQ	6	0	6
PCE	4	4	8
POE	4	2	6
OSCE	4	4	8

Discussion

Based on our findings, we developed the following models for selecting appropriate assessment methods for consideration for use in African medical schools.

The ideal model

The ideal model applies to any institution where fiscal and human resources are not so critically lacking as to limit the development of effective learning and assessment programmes. This is primarily the case at North American, United Kingdom, Western European and Australian medical schools. In this model, overall scores for the four major factors (performance, cost, suitability and safety) are weighted according to their importance at a particular institution and a total score is used to determine the most appropriate assessment method.

Table 5 illustrates application of the ideal model for selecting an assessment method. For this hypothetical institution, the ability to test for performance is the most important factor determining the choice of assessment method. Accordingly, performance is weighted by a factor of seven. Cost is weighted by a factor of two because, to a limited extent, it is a consideration for any medical institution. Suitability and examination safety issues are not major concerns, except for the need for continuous surveillance, so each is weighted by a factor of only 0.5.

In this ideal situation, OSCE are the most appropriate method of assessing student performance. This is true of current practice in developed world regions.[10-12] It should be noted that when institutions need to select a number of appropriate assessment methods. The model still serves as a useful guide in the selection process. The scores reflected in Table 5 indicate that PCE and POE would be preferable alternatives to OSCE rather that EQ or SAQ examinations. Based on the scores depicted in Table 5, MCQ examinations cannot be recommended for testing the domains of competence reflected in Table 1.

Table 5. *A weighed model for six assessment methods in an ideal context*

Assessment method	Performance (7)		Cost (2)		Suitability (0.5)		Safety (0.5)		Total weighted score
	Score	Score x wf	Score	Score x wf	Score	Score x wf	Score	Score x wf	
EQ	6	42	7	14	10	5.0	6	3	64.0
SAQ	6	42	7	14	10	5.0	6	3	64.0
MCQ	3	21	9	18	10	5.0	6	3	47.0
PCE	9	63	1	2	3	1.5	8	4	70.5
POE	8	56	4	8	8	4.0	6	3	71.0
OSCE	10	70	0	0	3	1.5	8	4	75.5

wf = weighting factor

Current African model

Table 6 illustrates the current situation at many medical institutions in Africa, and probably in other developing countries elsewhere. In Africa, limited human resources and fiscal constraints make cost a very important determinant of the choice of assessment methods used. In this model, we weighted cost by a factor of three. Suitability is also a cause for concern because many African medical institutions are currently admitting students beyond the capacity of the training facilities and personnel. As such, we weighted suitability by a factor of two. Examination safety problems, due to inadequate supervision and insufficient materials or facilities for examination administration, may also arise. Therefore, we weighted safety by a factor of one.

Table 6. *A weighted model for six assessment methods in the current African context*

Assessment Method	Performance (4)		Cost (3)		Suitability (2)		Safety (1)		Total weighted score
	Score	Score x wf	Score	Score x wf	Score	Score x wf	Score	Score x wf	
EQ	6	24	7	21	10	20	6	6	71
SAQ	6	24	7	21	10	20	6	6	71
MCQ	3	12	9	27	10	20	6	6	65
PCE	9	36	1	3	3	6	8	8	53
POE	8	32	4	12	8	16	6	6	66
OSCE	10	40	0	0	3	6	8	8	54

wf = weighting factor

Using this weighting strategy, EQ and SAQ examinations achieve the highest score and appear to be the most appropriate assessment methods given the extreme resource constraints. This reaffirms current student assessment practices widely prevalent in African medical schools, particularly in the pre-clinical years. Where possible, EQ and SAQ are supplemented by MCQ and/or oral examinations (vivas). In the clinical years PCE form part of the assessment process, where possible. Because patients constitute part of the clinical training programme, their use in PCE would not constitute an extra examination resource, except when they are paid.

An alternative African model

Because EQ, SAQ, and MCQ examinations are not optimal ways of assessing student performance, particularly skills – clinical, communication and problem-solving – institutions using these assessment methods cannot confidently claim to have achieved the objectives of most medical curricula, i.e. competent problem-solving clinicians.[7,8] The daunting task for African medical schools, therefore, is to search for assessment methods that suit local resource constraints without compromising educational standards. Table 7 illustrates an alternative model for selecting assessment methods at African medical institutions. This new model calls for substantive improvisation and restructuring of needs to minimize the influence of cost, suitability and safety on the selection of educationally appropriate assessment tools. Sharing or borrowing examination facilities, as well as the use of open wards or halls with shield curtains for patient-based examinations, could address many issues relevant to test suitability. The re-channeling of finances to where they are most needed, such as the recruitment of clinician-examiners and/or the purchase of essential examination materials would improve examination time and improve safety issues. Some of the essential examination materials, such as computers with the required software programmes or Internet capabilities, could further reduce examination preparation time. By implementing these measures, the influence of suitability and safety could be greatly reduced (perhaps to a weighting factor of 0.5 each). Cost, however, would remain a major factor impacting upon assessment selection and, therefore, would retain its weighting factor of three.

In the new model, the POE would be the most appropriate assessment method for an African medical institution. This is good news because this assessment technique addresses the important learning outcome of clinical reasoning skills. However, the implications of changing assessment methods in African medical schools must be considered before implementation. This is because assessment methods influence student learning behaviour, such that, a change in assessment methods would call for medical curriculum delivery changes.[12] We know that if assessment methods requires recall of facts, students tend to work on memorizing, whereas if examinations test problem-solving skills, students will focus on acquiring problem-solving skills.[13] Because the POE assessment would better suit students instructed in a problem-based

learning (PBL) programme, using this assessment method for students instructed in a traditional lecture-based programme could be considered unfair and may lead to failure. Therefore, adoption of the POE assessment method by African medical schools may require them to change their method of instruction to a PBL-type format, to a greater of lesser extent.

Table 7. *A weighted model for six assessment methods for use in an improved African context*

	Performance (6)		Cost (3)		Suitability (0.5)		Safety (0.5)		Total Weighted Score
Assessment Method	**Score**	**Score x wf**	**Score**	**Score x wf**	**Score**	**Score x wf**	**Score**	**Score x wf**	
EQ	6	36	7	21	10	5.0	6	3	65.0
SAQ	6	36	7	21	10	5.0	6	3	65.0
MCQ	3	18	9	27	10	5.0	6	3	53.0
PCE	9	54	1	3	3	1.5	8	4	62.5
POE	8	48	4	12	8	4.0	6	3	67.0
OSCE	10	60	0	0	3	1.5	8	4	65.5

wf = weighting factor

In this paper we have presented a framework by which to select appropriate assessment methods by systematically considering the factors that most significantly impact upon the selection of assessment methods, specifically in resource-constrained environments. The model is easy to understand and could be used as a tool for promoting change in both student assessment practices and medical curriculum delivery in African medical schools.

Acknowledgements

The authors thank the organizers of the third African course on pharmacotherapy training at Cape Town, (i.e., The World Health Organization and the Department of Pharmacology, University of Cape Town), for making this fruitful interaction of Africa's professionals happen, and the South Africa Drug Action Programme, for sponsoring the participants.

References

1. Kassebaum DG, Eaglen RH. Shortcomings in the evaluaton of students' clinical skills and behaviors in medical school. Academic Medicine 1999; 74: 842-849.
2. Watson RT, Romrell LJ. Mission-based budgeting: removing a graveyard. Academic Medicine 1999; 74: 627-640.
3. Mennin SP, Kalishman S. Student assessment. Academic Medicine 1998; 73 (Suppl.): S46-S54.
4. Stone SL, Qualters DM. Course-based assessment; implementing outcome assessment in medical education. Academic Medicine 1998; 73: 397-401.
5. Fowell SL, Bligh JG. Recent developments in assessing medical students. Postgraduate Medical Journal 1998; 74: 18-24.
6. Tervo RC, Dimitrievich E, Trujillo AL, Whittle K, Redinius P, Wellman L. Clerkship: an overview. South Dakota Journal of Medicine 1997; 50: 153-156.
7. Palmer D, Rideout E. Essays. In: Sharmon S, Norma G, editors. Evaluation methods: a resource book. Hamilton, Canada: McMaster University; 1995. p. 59-66.
8. Day SC, Norcini JJ, Diserens D, Cebul RD, Schwartz JS, Beck LH, et al. The validity of an essay test of clinical judgement. Academic Medicine 1990; 65: S39-S40.
9. Wakeford RE, Roberts S. Short answer questions in an undergraduate qualifying examination: a study of examiner variability. Medical Education 1984; 18: 168-173.
10. Singer PA, Robb A, Cohen R, Norman G, Turnbull J. Performance-based assessment of clinical ethics using an objective structured clinical examination. Academic Medicine 1996; 71: 495-498.
11. Des-Marchais JE, Vu NV. Developing and evaluating the student assessment system in the preclinical problem-based curriculum at Sherbrooke. Academic Medicine 1996; 71: 274-283.
12. Newble DI, Jaeger K. The effect of assessments and examinations on the learning of medical students. Medical Education 1983; 17: 25-31.
13. Hamilton JD. Establishing standards and measurement methods for medical education. Academic Medicine 1995; 70 (Suppl.): S51-S58.