

# AN AUDIT OF ASSESSMENT TOOLS IN A MEDICAL SCHOOL IN EASTERN SAUDI ARABIA

**Abdullah M. Al-Rubaish, MD,\* Khalid U. Al-Umran, MD, FRCPC,† Lade Wosornu, FRCS, FRCSE‡**

Departments of \*Internal Medicine, †Pediatrics and ‡Surgery, College of Medicine, King Faisal University, Dammam, Saudi Arabia

**مقدمة:** إن تطبيق المنهج الدراسي لا يمكن أن يكون متكاملًا بدون تطبيق أسلوب صحيح في التقييم، لذا ينبغي على المدرسين استعمال الطرق المناسبة في التقييم والتأكد من سلامتها من أجل الوصول إلى الأهداف التعليمية المنشودة.

**هدف الدراسة:** إيجاد قاعدة معلومات وتحليلها من أجل وضع توصيات مناسبة لتحقيق تطبيق المنهج بصورة صحيحة.

**موضع الدراسة:** الأقسام السريرية والأقسام قبل السريرية في كلية الطب – جامعة الملك فيصل – المملكة العربية السعودية.

**طريقة الدراسة:** وزعت استبانة للأقسام المعنية لمعرفة طرق تقييم الطلاب

**نتائج الدراسة:** تمت الاستجابة من قبل 31 من أصل 34 مادة دراسية تدرس في كلية الطب – جامعة الملك فيصل. استعملت الأسئلة ذات الخيارات المتعددة في كل مادة دراسية.

(31/28) إجابة واحدة صحيحة، (31/13) إجابة صح/خطأ – مجموعة من الأساتذة اختارت الأسئلة في أغلب الأحيان. استعملت (14/10) من الأقسام قبل السريرية أسئلة جديدة وكذلك (14/10) استعملت أسئلة مكررة، بينما اعتمدت الأقسام السريرية على الأسئلة المخزنة (17/16).

درجة النجاح 60% اعتمدت قبل الامتحان في أغلب الأحيان (31/28) – في ستة مواد دراسية استخدمت الأسئلة الطويلة وكان في خمسة منها توجد إجابة نموذجية، وفي مادة واحدة كان التصحيح من قبل أساتذتين كل على حدة. طريقة أوسكي استعملت (17/7) قسم سريري استخدمت العدسة الضوئية لتصحيح إجابات الاختبارات النظرية في ثلاثة مواد دراسية فقط.

تتم مراجعة للدرجات بعد الامتحان في (14/13) مادة دراسية قبل السريرية وفي (17/10) مادة دراسية سريرية. درجة صعوبة الامتحان وعامل التفريق استعمل في 4 مواد دراسية فقط من أصل 31 مادة.

الإفادة الراجعة استعملت في (14/12) مادة دراسية قبل السريرية وفي (17/7) مادة دراسية سريرية. قواعد الإمتحانات كانت متوفرة لدى (31/10) مشرف على المادة الدراسية.

التوصيات: تعتبر الأسئلة متعددة الإختيارات ذات الإجابة الصحيحة الواحدة أفضل طريقة لتقييم الناحية النظرية لدى الطالب خصوصاً إذا أعدت بطريقة علمية صحيحة لكن يجب أن تكون هنالك أسئلة جديدة لكل إمتحان، ويجب تحليل نتيجة الإمتحان بطريقة سليمة ومقارنتها بالنتائج السابقة. نقترح أن تستعمل طريقة أوسكي وطريقة أوسي بشكل واسع.

الأسئلة الطويلة النصية صح /خطأ ، املاً الفراغ والأسئلة متعددة الإختيارات ذات أكثر من إجابة صحيحة واحدة تعتبر غير مناسبة وينبغي عدم استعمالها

نقترح أن تقوم مجموعة من المدرسين بوضع الأسئلة واتخاذ القرارات بصورة جماعية ، الإفادة الراجعة للطلاب يجب أن يتم تفعيلها

أن وجود مركز متخصص للامتحانات يتبع لقسم التعليم الطبي يعتبر ضرورة 0 الحاجة ملحة لعمل دراسات مستقبلية لتقييم طريقة التقييم ومقارنتها بالكليات الصحية في المنطقة.

**مفتاح الكلمات:-** طريقة التقييم مراجعة المنهج الأسئلة متعددة الإختيارات 0

Correspondence to:

Prof. Khalid Al-Umran, P.O. Box 40140, Al-Khobar 31952, Saudi Arabia  
E-mail: khalid\_umran@yahoo.com or kalumran@yahoo.com

**Background:** Assessment has a powerful influence on curriculum delivery. Medical instructors must use tools which conform to educational principles, and audit them as part of curriculum review.

**Aim:** To generate information to support recommendations for improving curriculum delivery.

**Setting:** Pre-clinical and clinical departments in a College of Medicine, Saudi Arabia.

**Method:** A self-administered questionnaire was used in a cross-sectional survey to see if assessment tools being used met basic standards of validity, reliability and currency, and if feedback to students was adequate. Excluded were cost, feasibility and tool combinations.

**Results:** Thirty-one (out of 34) courses were evaluated. All 31 respondents used MCQs, especially one-best (28/31) and true/false (13/31). Groups of teachers selected test questions mostly. Pre-clinical departments sourced equally from "new" (10/14) and "used" (10/14) MCQs; clinical departments relied on 'banked' MCQs (16/17). Departments decided pass marks (28/31) and chose the College-set 60%; the timing was pre-examination in 13/17 clinical but post-examination in 5/14 pre-clinical departments. Of six essay users, five used model answers but only one did double marking. OSCE was used by 7/17 clinical departments; five provided checklist. Only 3/31 used optical reader. Post-marking review was done by 13/14 pre-clinical but 10/17 clinical departments. Difficulty and discriminating indices were determined by only 4/31 departments. Feedback was provided by 12/14 pre-clinical and 7/17 clinical departments. Only 10/31 course coordinators had copies of examination regulations.

**Recommendations:** MCQ with single-best answer, if properly constructed and adequately critiqued, is the preferred tool for assessing theory domain. However, there should be fresh questions, item analyses, comparisons with previous results, optical reader systems and double marking. Departments should use OSCE or OSPE more often. Long essays, true/false, fill-in-the-blank-spaces and more-than-one-correct-answer can be safely abolished. Departments or teams should set test papers and collectively take decisions. Feedback rates should be improved. A Center of Medical Education, including an Examination Center is required. Fruitful future studies can be repeat audit, use of "negative questions" and the number of MCQs per test paper. Comparative audit involving other regional medical schools may be of general interest.

**Key Words:** Assessment Technique, Curriculum review, MCQ

---

## INTRODUCTION

Assessment is a critical component of instruction. Properly used, it can help institutions attain key curricular objectives. The aims of assessment include communicating what institutions see as important, and helping them monitor the program by providing feedback on the extent to which teaching results in learning. It can also disclose instructional gaps and encourage learners to read broadly and participate actively as and when educational opportunities become available.<sup>1</sup>

The impact of decisions on how and when to evaluate learners' knowledge and competence cannot be overstated. Tests are a powerful motivator, and learners tend to study what they believe instructors value. ("Assessment drives learning.") Thomas Huxley was quoted as saying: "Students work to pass, not to know. They do pass but they do not know."<sup>2</sup>

Because assessment has a powerful influence on all key aspects of learning and curriculum delivery, tools that reinforce educational goals should be used. Medical instructors must "ensure that assessment tools and their use conform to principles and procedures of educational science, and seek to improve the tools they intend to use."<sup>2</sup>

Thus, an audit of assessment tools should be an integral part of formal curriculum review. Furthermore, the last decade has seen an evolution of assessment tools in medical education from the traditional ones to more sophisticated tools such as OSCE, the portfolio approach and hi-tech simulations. It is necessary to be methodical in program evaluation which includes re-examining existing tools.<sup>3</sup>

We have performed one aspect of formative program evaluation, namely, an audit on the assessment tools being used. It was basically a quality assessment exercise, an audit of structure

defined as a survey to count, see and show whether acceptable standards are being met. No other value judgment was required.

The desirable attributes of assessment tools audited in this survey were validity, reliability, effect on students and whether they are up to date. Excluded were cost and feasibility of administering the tool, as well as strategies of employing two or more tools to assess the same course. It is expected that as our primary aim, the findings will provide valid information and support recommendations to improve the delivery of our curriculum. The secondary aim is an eventual improvement in the performance of learners and the program.

## SUBJECTS AND METHODS

A prospective survey was conducted in the College of Medicine, King Faisal University, Eastern Saudi Arabia in May 2004. A self-administered questionnaire was the survey tool used. Copies of the questionnaire were distributed to all course coordinators in the College, with the exception of English Language, Islamic Studies and Physical Education. Respondents were allowed to complete the questionnaire without prompting. Data were analyzed by SPSS software program to determine basic frequency distribution of the assessment tools being currently used.

Although it covered both Basic Sciences and Clinical streams in the College, the survey was not designed to yield data for statistical comparison. Hence, no statistical analysis was indicated and none was attempted.

## RESULTS

A total of 31 courses were evaluated. The type and subtype of written questions used were explored in questions 2 & 3 (Table 1). More than one response was allowed. MCQs emerged as the commonest; it was used by all 31 respondents. Long essays, short essay questions and fill-in-the-blank-spaces were infrequently used, but, short notes were employed by 9/14 pre-clinical departments. "One-best" response was the most frequently used subtype, and the least often were "extended matching items" and "more than one correct answer". However, "true/false" response was used by 7/14 pre-clinical and 6/17 clinical departments.

**Table 1:** Distribution of responses to types and sub-types of questions used

Question type	Department		
	Pre-clinical	Clinical	Both
Multiple choice questions	14	17	31
One best response	12	16	28
True/False response	7	9	16
Extended matching items	2	3	5
More than 1 correct answer	1	1	2
Patient Management Problem	3	2	5
Short Notes	9	3	12
Multiple Essay Questions	2	1	3
Long Essay	1	2	3
Fill in the blanks	2	0	2

Questions 4 & 5 dealt with the source of the selected questions (where and who). The group of teachers responsible for the course was the body which selected test questions most often, with the whole department a close second in clinical departments (10/17), but not in pre-clinical ones where individual teachers were used just as often (3/14). Whereas pre-clinical departments sourced equally from "new" (10/14) and "used" (10/14) MCQs, clinical departments relied heavily on 'banked' ones (16/17).

Questions 6-8 dealt with decisions about pass marks: who, how and when. Departments mostly took the decision: 12/14 pre-clinical and 16/17 clinical. The 60% pre-set by the College was the most frequently chosen pass mark: 10/14 and 15/17 for pre-clinical and clinical departments respectively. Whereas the decision was taken before examinations by most clinical departments (13/17), 5/14 pre-clinical ones did so after the examination.

How test questions were marked was addressed in questions 9-11 (Table 2). Of the 6 essay users in pre-clinical departments, 5 provided model answers but only one used double marking. OSCE was used by only of seven clinical departments, five of which provided check list. Optical reader was used in only 1/14 pre-clinical and 2/17 clinical departments.

Question 12 asked whether examinations were reviewed after marking, and if so, which review activities were used. Of 14 pre-clinical departments, 13 reviewed examinations post-marking. However, only 10 of 17 clinical departments did so. The types of review activities were as follows: difficulty and discriminating indices were determined by only four

departments; seven pre-clinical departments compared scores with previous years; seven with those in the same year. However, only five clinical departments performed this review activity.

Table 2: Distribution of responses on how questions were marked

Questions	Department		
	Pre-clinical	Clinical	Both
Q.9: Essays			
Same person marks same questions	6	1	7
Model answer available	5	2	7
Double marking	1	1	2
Q.10: OSCE			
Check list	-	5	5
Individual teacher	1	2	3
Q.11: MCQ			
Key used	8	9	17
Double check	10	5	15
Optical reader	1	2	3

OSCE=Objective structured clinical examination  
MCQ= Multiple choice questions

The form and timing of feedback during in-course assessment was explored in question 13. Of the 17 clinical departments, only seven provided feedback as against 12 in 14 pre-clinical ones where the commonest form was "exams discussed without exam papers" (5/12). The most frequent timing was within one week: 7/12 for pre-clinical and 5/7 clinical departments respectively.

Questions 14-16 explored three aspects of the conduct of examinations. Asked if students had prior access to examination papers, the responses were as follows: pre-clinical departments 13/14 'no' and one "yes"; clinical departments 16/17 'no' and one "don't know". As to whether students were appropriately informed about the mechanics of the conduct of the examination, all departments replied "yes". However, only 3/14 pre-clinical and 7/17 clinical course coordinators had copies of the booklet on examination regulations.

## DISCUSSION

The population studied was representative of the College faculty: all except three course coordinators in basic sciences and clinical departments were surveyed. Rating forms, questionnaire and performance audits are the

methods of measurement used to evaluate educational programs.<sup>3</sup> Self-administered questionnaire was the method used here. As such, it was considered adequate since the study was a basic formative program assessment, not a summative one which would seek to judge performance.

It was gratifying that MCQs were used by all respondents and that the one-best subtype was the most frequent. However, observed deficiencies include lack of item analyses and regular up-dates of MCQ bank. Some departments failed to provide answer keys or use double marking. The use of optical reader systems was negligible.

It was also encouraging that infrequently used question types were fill-in-the-blank-spaces, extended matching items and more than one correct answer. OSCE was used by seven clinical departments and five provided check lists.

However, it was disturbing that in this study, long essays were still being used especially by some pre-clinical departments. This deficiency was aggravated by the non-use of double marking. Medical educationists agree that, as an assessment tool, the long essay is out-of-date.<sup>4</sup> Paul observed: "Long essay questions have limited reliability and poor validity. They are not an objective measure of learning outcome. They have little role in medical education."<sup>4</sup>

The observed frequent use of true/false subtype was equally disturbing because of its known flaws.<sup>1</sup> It may be easier to construct than one-best format, but it is more problematic. The student is guaranteed 50% chance of guessing the correct answer. Though the original item writer had a particular fact in mind when he wrote the question, it can be ambiguous, or the distinction between "true" and "false" blurred and obscure. Thus, subsequent reviewers alter the answer key, rewrite or discard the question more frequently than items written in other MCQ formats. Whereas some ambiguities can be clarified, others cannot. One way to avoid ambiguity is to test for simple recall of isolated facts, although educationists discourage this practice.<sup>1</sup>

It was appropriate and commendable that test questions were most often selected by the group of teachers responsible for the course, or the department as a whole. The use of MCQ implies that the team responsible for the course should be involved since it is unlikely that one individual

can develop a bank of well evaluated MCQs.<sup>5</sup> Whereas pre-clinical departments correctly sourced in equal measure from new and used MCQs, the observed heavy reliance on banked MCQs by clinical departments was inappropriate.

Departments took decisions on pass marks and chose the 60% pre-set by the College; this was appropriate. However, whereas the decision was taken before examinations by most clinical departments, it was disturbing that many pre-clinical departments did so after examination.

The frequency of post-examinations review was satisfactory in pre-clinical but not in clinical departments. However, throughout the College, basic item analyses such as calculating difficulty and discriminating indices were grossly deficient. Similarly, comparing scores with previous years or the same year was infrequently practised, especially in clinical departments.

The low feedback rate observed in clinical departments was unsatisfactory. Learners require regular feed back on what they know or do not know in order to learn from their mistakes.<sup>1-3</sup> Assessment also affects students' self-esteem, career aspirations and accomplishments.<sup>5</sup> The provision of feedback within one week was a pragmatic approach and can be encouraged. Course coordinators are urged to obtain copies of the booklet on examination regulations.

## CONCLUSIONS AND RECOMMENDATIONS

No one assessment tool by itself is perfect, and the "pivotal role of assessment in the educational process"<sup>7</sup> cannot be over-emphasized. After a careful review of the literature, MCQ with single-best answer, if well constructed and adequately critiqued, emerged as the preferred tool for assessing the theory domain. Higher taxonomies such as application of knowledge, integration, synthesis and judgment can also be assessed by it if based on patient management problems. However, it should be enhanced by the following means, among others.

Faculty members in general and course coordinators in particular should become familiar with item analyses and their correct application, as well as appropriate comparisons with previous results. The College should provide optical reader

systems to all departments. Until departments have adequate banks of MCQs, each test paper should contain at least 50% new questions. All departments should provide answer keys, and practise double marking.

Departments should use OSCE or OSPE more often along with check lists. Four question types can all be abolished without educational loss to the Curriculum: long essays, true/false, fill-in-the-blank-spaces and more than-one-correct-answer.

At all times, the department as a whole or the team of instructors responsible for the course should set the test paper and collectively take all decisions, including the pass mark. On no account should such matters be left to one individual to finalize. Clinical departments should improve on feedback rates at in-course assessments. A Center of Medical Education, including an Examination Center, requires to be established as soon as possible. This will, among other benefits, permit the College to play an informed role in "BEME"-Best Evidence Medical Education.<sup>6,7</sup>

It remains to be seen the extent to which these recommendations will be implemented. In order to complete the audit cycle, a repeat audit is mandatory. Furthermore, aspects of assessment tools which can be explored in other studies include "negative questions" and the number of MCQs per test paper. Finally, a comparative audit involving other regional medical schools may be of general interest.

## REFERENCES

1. Al Umran K. "Format for Written Examinations." Personal Communication. July 2003.
2. Abeykoon P. "Foreword " K. L. Wig Centre for Medical Education and Technology, All India Institute of Medical Sciences. SEARO, WHO Project I—R / IND HRH 001/LGS 1995
3. Wojtczak A. "Evaluation of Learning Outcomes." [http:// www.iime.org/documents/elo.htm](http://www.iime.org/documents/elo.htm). 7-8-2003
4. Paul V. K. "Essay Questions", in "Assessment in Medical Education: Trends and Tools." K. L. Wig Centre for Medical Education and Technology, All India Institute of Medical Sciences. SEARO, WHO Project I—R / IND HRH 001/LGS 1995
5. DAMMCQ: Designing and Managing MCQs – [http:// web.uct.ac.za/projects/cbe/mcqman/mcqch2.html](http://web.uct.ac.za/projects/cbe/mcqman/mcqch2.html).
6. Best Evidence Medical Education. Report of a Meeting. *Medical Teacher* 2000; 22: 242-5.
7. Harden RN, Davis MH, Friedman BD. UK recommendations on undergraduate medical education and the Flying Wallendas. *Med Teacher* 2002;24:5-8