Joshua Freeman, MD, Feature Editor
Alison Dobbie, MD, Associate Editor

Editor's Note: Send submissions to jfreeman@kumc.edu. Articles should be between 500–1,000 words and clearly and concisely present the goal of the program, the design of the intervention and evaluation plan, the description of the program as implemented, results of evaluation, and conclusion. Each submission should be accompanied by a 100-word abstract. Please limit tables or figures to one each. You can also contact me at Department of Family Medicine, KUMC, Room 1130A Delp, Mail Code 4010, 3901 Rainbow Boulevard, Kansas City, KS 66160. 913-588-1944. Fax: 913-588-2496.

Reliability and Validity Testing of an Evidence-based Medicine OSCE Station

Fred Tudiver, MD; Doug Rose, MD; Burt Banks, MD; Deborah Pfortmiller, MA

The six competencies of the Accreditation Council for Graduate Medical Education include the lifelong learning skills of evidence-based medicine (EBM)/information mastery. We developed and tested an Objective Structured Clinical Examination (OSCE) station that would measure these skills in family medicine residents. This EBM OSCE station is a 30-minute station within a regular OSCE exam. It uses an 8-point checklist and global measure and has good psychometric properties, including construct validity, interrater reliability (correlation=.96), and internal reliability (Cronbach's=.58). This tool is useful for training programs, since assessing EBM/information mastery is an important part of the evaluation of physician skills.

(Fam Med 2009;41(2):89-91.)

In 1999, the Outcomes Project of the Accreditation Council for Graduate Medical Education (ACGME) developed six core competencies—all to be implemented by July 2007. The competency areas of medical knowledge and practice-based learning and improvement specifically included the concepts of lifelong learning. The second phase of the project (2002–2006) asked training programs to assess the six competencies, with the aid of a "toolbox" of suggested assessment techniques.[1] For the lifelong learning competencies, the toolbox often included the use of Objective Structured Clinical Examinations (OSCEs).

The specific lifelong learning competencies overlap with the four major skills of evidence-based medicine (EBM): (1) Translation of uncertainty to an answerable question, (2) Systematic retrieval of best evidence available, (3) Critical appraisal of evidence for validity, clinical relevance, and applicability, and (4) Application of results in practice.[2]

A recently published systematic review of 104 unique instruments for evaluating the teaching of EBM skills found most were tested on medical students and residents, and most were restricted to assessing skills in searching and critical appraisal.[3] The reviewers concluded that of the 104 instruments, 34 measured actual EBM clinical behaviors. Of these, only six used objective outcome measures, and only three measured the performance of evidence-based clinical maneuvers in practice.[4-6] The rest relied on retrospective self reports. These three measures recorded practice audits as a proxy for visualizing actual practice itself.

Another four studies in the review used OSCEs to assess EBM skills on medical students, but they had limitations. One was restricted

From the Department of Family Medicine, East Tennessee State University.

to assessing searching strategies,[7] another did not assess searching or critical appraisal skills,[8] and a third was restricted to assessing critical appraisal.[9] The fourth study was the only study of the four that examined psychometric properties of the measure, but it did not assess how the searches were performed.[10] A paper on a fifth OSCE-related EBM skills measure (also with medical students) was published after the systematic review, but it did not appraise search skills.[11] Most of the OSCE-based studies restricted their searching databases to MEDLINE.

The purpose of this pilot study was to develop and test an OSCE station that measured EBM skills.

Methods

Setting and Subjects

Three of the authors initially developed a set of two EBM OSCE stations based on two different integrated OSCEs that are given to the incoming first- and second-year residents in the Department of Family Medicine at East Tennessee State University as a formative evaluation in the first month of their year. Twenty-three first-year residents and 19 second-year residents completed the testing. The integrated OSCE consisted of six different stations (standardized patient interview, focused clinical examination, interpretation of lab findings, development of differential diagnosis and plan, exploration of ethics, and confidentiality issues), with an innovative EBM station at the end. The two OSCEs were based on two cases: (1) For first-year residents, a patient with multiple myeloma presenting with back pain, and (2) For second-year residents, a patient was admitted with pancreatitis and alcohol abuse.

The EBM OSCE Stations

The residents were given 30 minutes to complete three sections of inquiry for each of the OSCEs. In the first, they had to develop a four-part P.I.C.O. question (Patient-Intervention-Comparison if any-Outcomes) related to the OSCE case. In the second section, they were given a P.I.C.O. question based on the same case and asked to find the best evidence answer. They were provided with a computer connected to the Internet and all the medical college library resources, including MEDLINE, Cochrane Databases of Systematic Reviews and Clinical Trials, D.A.R.E., ACP Journal Club, InfoRetriever, InfoPoems, UpToDate, and DynaMed. Residents completed a form documenting the resources searched, terms used, type of studies found, usefulness, best evidence answer to the question, and justification for choosing the study or studies. The third section contained seven multiple-choice questions assessing comprehension of levels of evidence and understanding of disease-oriented evidence (DOE) versus Patient-oriented Evidence That Matters (POEM).[12]

Testing and Scoring

Most modern OSCEs utilize both a checklist and a global scoring mechanism.[13,14] We devised a similar set of evaluation measures for the EBM OSCE, with response scale items based on the EBM literature. After three iterations of testing we came up with a revised eight-item checklist. The eight items in the checklist included questions on the four major EBM skills as well as on efficiency finding answers to their questions and assessing levels of evidence for critically appraising articles. The possible scores of the eight-item checklist ranged from zero to 24 and were computed by summing the eight items, each with a response range of zero to 3. The two-item global scale (one item for process, the second for the answer) ranged from zero (no response) to 10 (highly effective and efficient search). Three author raters (FT, DR, BB), independently scored the checklists and global scores after a number of discussions and agreements on how to score.

Data Analysis

We measured content validity by using feedback from expert opinion and construct validity by using independent *t* tests to compare the means of the first- and second-year residents. Pearson correlation coefficient was used to examine the strength of the relationship between the checklist measure and the global measure, giving a measure of criterion validity. We measured interrater reliability by using two-way mixed effects intraclass correlations for consistency and internal reliability by using Cronbach's alpha.

Results

There was good construct validity as new rising second-year residents (most of whom had EBM training) had higher scores than new first-year residents on the global assessment score (Table 1). Second-year residents also scored higher on the eight-item checklist score, but this difference was not significant (PGY-1 mean=15.05, PGY-2 mean=16.37). The checklist and global assessment measure had a statistically significant positive correlation ($r=0.62$, $P<.001$). The final eight-item checklist and the global assessment had good interrater reliability (0.96 and 0.92 respectively) (Table 1). The internal reliability of the eight-item scale as measured by Cronbach's alpha was 0.58, considered acceptable.[15]

Discussion/Conclusions

This brief report describes the development and testing of a new, innovative assessment tool for evaluating the four major skills of EBM. In addition, it tests these skills in a simulated clinical situation—the OSCE. The tool is flexible and can be used in almost any clinical OSCE situation, once the OSCE itself is developed.

Table 1

Properties and Results of the EBM OSCE Station

Test Property	*Measure Utilized*	*Results*
Content validity	Expert opinion	Covered critical EBM skills revisions based on feedback
Construct validity (ability of an instrument to measure an abstract concept)	Mean scores of first year compared to second year	First-year mean global scores lower than second-year (5.65 versus 6.95), t=-2.10, P=.043
Criterion validity (ability of one test to predict the results obtained in another test)	Global measure correlated with checklist measure	Good agreement (0.62, P<.001)
Interrater reliability (degree of agreement in scores among three raters)	Interrater correlation for checklist and global measures	High agreement 0.96 for checklist 0.92 for global score
Internal reliability (degree to which groups of test questions measure a single construct)	Cronbach's alpha	Acceptable (0.58)

EBM—evidence-based medicine
OSCE—objective structured clinical exam

Some of the psychometric properties (construct validity, interrater reliability, criterion validity) were good, including the criterion validity since the global and checklist correlation was highly significant; however, although the Cronbach's alpha reliability of 0.58 is considered acceptable to evaluate level of group accomplishment,[15] other newer references indicate that 0.70 is a standard level.[16,17] This, and the lack of a significant difference in the checklist between the two resident years, could be a problem of power with the relatively low numbers of residents in the study.

The development of standardized tools for assessing skills of EBM/information mastery is becoming an essential part of the evaluation of physician skills. Until we have integrated these tools routinely in our education of these physicians and other health professionals, we will be unable to evaluate performance in one of the key competency skills—that of lifelong learning. Future development of this tool will need testing on larger numbers and more rigorous testing of the psychometric properties.

Acknowledgments: Portions of the content of this paper were presented at the 2004 Society of Teachers of Family Medicine Annual Spring Conference in Toronto.

Corresponding Author: Address correspondence to Dr Tudiver, East Tennessee State University, Department of Family Medicine, Box 70621, Johnson City, TN 37614. 423-439-6738. Fax: 423-439-2440. tudiverf@etsu.edu.

References

1. Accreditation Council for Graduate Medical Education. Table of toolbox methods. www.acgme.org/Outcome/assess/ToolTable.pdf. Accessed February 12, 2008.
2. Dawes M, Summerskill M, Glasziou G, et al. Sicily statement on evidence-based practice. BMC Med Educ 2005;5(1):1.
3. Shaneyfelt T, Baum KD, Bell D, et al. Instruments for evaluating education in evidence-based practice. JAMA 2006;296:1116-27.
4. Langham J, Tucker H, Sloan D, Pettifer J, Thom S, Hemingway H. Secondary prevention of cardiovascular disease: a randomised trial of training in information management, evidence-based medicine, both or neither: the PIER Trial. Br J Gen Pract 2002;52:818-24.
5. Epling J, Smucny J, Patil A, Tudiver F. Teaching evidence-based medicine skills through a residency developed guideline. Fam Med 2002;34(8):646-8.
6. Ellis J, Mulligan I, Rowe J, Sackett DL. Inpatient general medicine is evidence based: A-Team, Nuffield Department of Clinical Medicine. Lancet 1995;346:407-10.
7. Burrows SC, Tylman V. Evaluating medical student searches of MEDLINE for evidence-based information: process and application of results. Bull Med Libr Assoc 1999;87:471-6.
8. Fliegel JE, Frohna JG, Mangrulkar RS. A computer-based OSCE station to measure competence in evidence-based medicine skills in medical students. Acad Med 2002;77:1157-8.
9. Bradley P, Humphris G. Assessing the ability of medical students to apply evidence in practice: the potential of the OSCE. Med Educ 1999;33:815-7.
10. Davidson RA, Duerson M, Romrell L, Pauly R, Watson RT. Evaluating evidence-based medicine skills during a performance-based examination. Acad Med 2004;79:272-5.
11. Frohna JG, Gruppen LD, Fliegel JE, Mangrulkar RS. Development of an evaluation of medical student competence in evidence-based medicine using a computer-based OSCE station. Teach Learn Med 2006;18:267-72.
12. Ebell MH, Shaughnessy A. Information mastery: integrating continuing medical education with the information needs of clinicians. J Contin Educ Health Prof 2003;23(Suppl 1):S53-S62.
13. Jefferies A, Simmons B, Tabak D, et al. Using an objective structured clinical examination (OSCE) to assess multiple physician competencies in postgraduate training. Med Teach 2007;29:183-91.
14. Berkenstadt H, Ziv A, Gafni N, Sidi A. The validation process of incorporating simulation-based accreditation into the anesthesiology Israeli National Board Exams. Isr Med Assoc J 2006;8:728-33.
15. Helmstadter GC. Principles of psychological measurement. New York: Appleton Century Crofts, 1964.
16. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use, second edition. New York: Oxford University Press, 1995.
17. Polit DF, Beck CT. Nursing research: principles and methods, seventh edition. Philadelphia: Lippincott Williams & Wilkins, 2004.