

Learner perception of oral and written examinations in an international medical training program

Sean P. Kelly · Scott G. Weiner · Philip D. Anderson ·
Julie Irish · Greg Ciottone · Riccardo Pini ·
Stefano Grifoni · Peter Rosen · Kevin M. Ban

Received: 15 July 2009 / Accepted: 23 November 2009 / Published online: 5 February 2010
© Springer-Verlag London Ltd 2010

Abstract

Background There are an increasing number of training programs in emergency medicine involving different countries or cultures. Many examination types, both oral and written, have been validated as useful assessment tools around

Electronic supplementary material The online version of this article (doi:10.1007/s12245-009-0147-2) contains supplementary material, which is available to authorized users.

The views expressed in this paper are those of the author(s) and not those of the editors, editorial board or publisher.

S. P. Kelly (✉) · P. D. Anderson · G. Ciottone · P. Rosen ·
K. M. Ban

Department of Emergency Medicine, West CC-2,
Beth Israel Deaconess Medical Center,
300 Brookline Avenue,
Boston, MA 02215, USA
e-mail: skelly2@bidmc.harvard.edu

S. G. Weiner
Department of Emergency Medicine, Tufts Medical Center,
Boston, MA 02111, USA

J. Irish
Shapiro Center for Education,
Beth Israel Deaconess Medical Center,
Boston, MA, USA

R. Pini
Department of Critical Care Medicine and Surgery,
University of Florence and Azienda
Ospedaliero–Universitaria Careggi,
Florence, Italy
e-mail: rpini@unifi.it

S. Grifoni
Department of Emergency Medicine,
Azienda Ospedaliero–Universitaria Careggi,
Florence, Italy

the world; but learner perception of their use in the setting of cross-cultural training programs has not been described.

Aims The goal of this study was to evaluate learner perception of four common examination methods in an international educational curriculum in emergency medicine.

Methods Twenty-four physicians in a cross-cultural training program were surveyed to determine learner perception of four different examination methods: structured oral case simulations, multiple-choice tests, semi-structured oral examinations, and essay tests. We also describe techniques used and barriers faced.

Results There was a 100% response rate. Learners reported that all testing methods were useful in measuring knowledge and clinical ability and should be used for accreditation and future training programs. They rated oral examinations as significantly more useful than written in measuring clinical abilities ($p < 0.01$). Compared to the other three types of examinations, learners ranked oral case simulations as the most useful examination method for assessing learners' fund of knowledge and clinical ability ($p < 0.01$).

Conclusions Physician learners in a cross-cultural, international training program perceive all four written and oral examination methods as useful, but rate structured oral case simulations as the most useful method for assessing fund of knowledge and clinical ability.

Keywords Graduate medical education · Oral case simulations · Assessment tools · Curriculum development · International medical education

Introduction

Medical educators around the world have successfully used many different methods of assessing learners, both written

and oral [1]. Multiple-choice and essay examinations have been a mainstay at every level of medical education in many countries. Additionally, there is a growing body of evidence that oral examinations, including case simulations in particular, can be important assessment tools in medical education [2–4].

Medical training programs in many countries use oral case simulations as assessment tools [5]. Many recognized clinical skills training programs such as basic life support (BLS), advanced cardiac life support (ACLS), and pediatric advanced life support (PALS) employ case simulations for teaching and assessment [6–8]. In emergency medicine (EM) and other specialties, oral case simulations are used extensively for teaching, assessment [9, 10], and certification [11–14].

Despite the evidence that the use of these and other methods leads to effective learner assessment in various countries [15–17], there has been little published on their use in cross-cultural, international medical training programs [18, 19]. Particularly with oral examinations, the question arises as to whether they can be useful in international programs in which teachers and learners may encounter language barriers or cultural differences.

In this paper we describe the learner perception of four common methods of testing (multiple-choice tests, essay tests, structured oral case simulations, and semi-structured oral examinations) used as part of the needs assessment (pre-testing) and qualification process (post-testing) in an international EM training program in Tuscany, Italy [20]. We also describe the techniques used and barriers faced in the examination process.

Objective

The aim of the study was to evaluate learner perception of four common examination methods in an international educational curriculum in EM: structured oral case simulations, multiple-choice tests, semi-structured oral tests, and essay tests.

Methods

Study design

This was a prospective, observational study using an assessment tool to evaluate learner satisfaction with four different examination methods used in a cross-cultural medical training program. This study was approved for exemption by the Institutional Review Board of the Azienda Ospedaliero-Universitaria Careggi, which is the University Hospital in Florence, Italy.

Study setting and population

The Tuscan Emergency Medicine Initiative (TEMI) is an international partnership involving the Tuscan Ministry of Health, the Tuscan University system, Harvard Medical International (HMI), and the Beth Israel Deaconess Medical Center (BIDMC) Department of Emergency Medicine in Boston, MA, USA. Its goal is to develop an EM training infrastructure for physicians working in the regional hospital system [21]. At the onset, 24 practicing Italian physicians participated in an EM train-the-trainers program based at the University Hospital in Florence, Italy from June 2003 to April 2004. Prior to the start of the program, participants were given written and oral pre-tests in order to evaluate their knowledge base in EM. At the end of the program they were given written and oral post-tests for summative assessment and qualification as EM educators in the region.

Examination methods

The oral case simulations were a pre-test which was used as part of the needs assessment for the project [22]. Ten structured oral case simulations based on clinical scenarios were prepared in a uniform format, with history, physical examination, radiological studies, laboratory results, and visual stimuli available when appropriate. The scenarios and questions asked were scripted in a uniform manner and there were critical actions which needed to be performed by examinees. All written materials were available in Italian and testing was conducted with a medical translator who, in addition to being fluent in Italian and English, was also an emergency physician and a content expert in the subject matter. Candidates were expected to complete three cases chosen randomly from the ten prepared cases. One case was selected for content with which they had adequate prior postgraduate training (internal medicine). Two cases were selected for content with which they had minimal postgraduate training (trauma, surgery, ophthalmology, wound care, etc.). Please see ESM Figure 1 for an example of one of the cases used. Twenty minutes were allotted for each case. For each test, one examiner administered the test and the other observed. The scores from both the examiner and the observer were used for the final score.

The multiple-choice examination was a written pre-test composed of 75 multiple-choice questions selected from various test preparation materials used in the USA and Europe and modified to cover the intended curriculum [22]. The questions were translated into Italian and edited by an Italian clinician for accuracy and local clinical relevance.

The oral semi-structured examination was a post-test that was similar to the oral pre-test, but was not as rigidly structured. The same basic format and testing procedures

were used except for the following: Since these examinations were used for qualification purposes, highly experienced examiners not directly affiliated with our training program were brought in as experts to examine the participants. The beginning of each case was structured in a similar fashion as the pre-test with scripted clinical scenarios and prepared materials, but the examiners were allowed more flexibility to ask unstructured follow-up questions to assess elements of the examinees' fund of knowledge, points of management, and decision-making logic.

The essay test was a written post-test that was composed of four short answer essay questions. Examinees were informed of the general topics to be covered ahead of time (which consisted of the major topics covered in the curriculum during the training program), but the actual specific questions were unknown to the examinees until the time of the test. The test was graded according to whether they addressed the major critical topics correctly.

A learner satisfaction survey was obtained at the end of the training program that asked the Italian physicians to rank the four different examination methods in order of preference according to "usefulness in assessing fund of knowledge" and "usefulness in assessing clinical abilities." They were also asked to rate the oral and the written examination difficulty levels on a 1–5 Likert scale (anchors of 1 = extremely difficult, 2 = too difficult, 3 = appropriate, 4 = too easy, 5 = extremely easy). They were asked if the written and oral examinations were useful in measuring fund of knowledge and clinical ability (yes/no answers) and whether they should be used in future programs or for accreditation to practice EM in their region (yes/no answers). This was a written survey that was conducted as part of the end-of-the-year course evaluation (ESM Figure 2). Participants were asked if they would give honest feedback to help improve the process for future learners. As a result, the physician learners were blinded to the purpose of the learner satisfaction survey and all responses were anonymous.

Statistical analysis

Because the study outcomes were not normally distributed, comparisons were made using the following nonparametric tests: the Wilcoxon rank sum test was used to compare the difference in the mean content difficulty ratings of oral versus written examinations; the Fisher exact test (due to the small samples) was used to compare survey responses that were in yes/no format; and the Friedman repeated measures test was used to compare mean rankings of examination usefulness in cases when there was one categorical independent variable (examination type) and one continuous variable (mean rank score for fund of

knowledge and clinical abilities). All statistical analyses were performed using SPSS version 14.0 (Chicago, IL, USA).

Results

There was a 100% response rate. All 24 participants responded to the survey, but there were data missing from 2 participants (6 answers total). All areas of missing data are noted in the text or tables.

The respondents found the oral examinations slightly more difficult than the written examinations. The mean difficulty rating was 2.75 for oral examinations and 3.00 for written examinations, with a standard deviation of ± 0.45 ($Z=-2.45$; $p<0.01$). There was one respondent who did not answer the question on the difficulty of the oral examinations.

In general, learners liked all testing methods, with the majority of learners responding in the affirmative when asked whether each examination method was valuable for use in future programs and accreditation, and for measuring fund of knowledge or clinical abilities. Only the perceived usefulness in measuring clinical abilities was found to be significantly higher in oral (83%) versus written (67%) examinations ($p<0.01$). Please see Table 1.

Using the Friedman repeated measures test, we found a significant difference among the four types of examinations for assessing the fund of knowledge (chi-square 16.42, $p<0.001$) and clinical abilities (chi-square 14.23, $p<0.01$). Post hoc Wilcoxon (Bonferroni corrected) indicated that the structured oral pre-test was ranked significantly higher than the other three examinations on both measures: fund of knowledge ($p<0.01$) and clinical abilities ($p<0.01$). No other pairwise comparisons among the other three types of examinations were found to be significant. Please see Table 2.

Discussion

Learner assessment is a complex process and there are various methods which can be used [1]. Each of the methods described in the literature have their own strengths and weaknesses [15]. Many authors believe that the use of multiple methods of assessment in any one training program can overcome the limitations of individual methods and enhance the overall validity and effectiveness of learner assessment [16, 17]. Furthermore, different methods may be more effective at assessing the different levels of Miller's framework of clinical assessment [23]. Structured case simulations may provide educators with a better assessment of a learner's behavior (and therefore

Table 1 Learner perception of examination usefulness

	Oral examinations	Written examinations
Should this examination type be used in training programs in the future?	96% yes	96% yes
Should this examination type be used for accreditation in EM?	100% yes	92% yes
Useful in measuring fund of knowledge?	100% yes	96% yes
Useful in measuring clinical abilities? ^a	83% yes ^b	67% yes

^a Statistically significant difference between oral and written testing methods in measuring clinical abilities (chi-square=11.07, $p<0.01$). No other statistically significant differences noted between oral and written testing methods

^b Missing data: 1 of the 24 respondents did not respond to this question

predicted clinical performance), rather than simply his cognition [1, 23, 24]. It is important to distinguish the structured and semi-structured nature of the oral examinations we used in this training program from the traditional unstructured oral examinations used in the past in many places, including Italy. Most authors agree that structured examinations have better validity and reliability, with less susceptibility to gender or cultural bias than unstructured examinations [20, 25, 26].

There are multiple examples of training programs in countries throughout the world using case-based oral pre-tests for needs assessment and oral post-tests for assessment of learners or accreditation of physicians [27, 28]. However, there are few descriptions of cross-cultural international educational programs utilizing these same methods. We found oral pre- and post-tests extremely valuable to educators and well-received by learners in this training program and offer our perspective in the hopes that others will be encouraged to incorporate this methodology into similar ventures. International medical education presents a host of unique opportunities: the opportunity to learn, to teach, and to share knowledge beyond preconceived boundaries or borders. With these opportunities come unique challenges: the tasks of bridging language, cultural, knowledge, and experiential differences [29].

In considering the question of whether case-based oral testing can be used in an international training program, we found several barriers that needed to be addressed. The potential language barriers were mitigated by having all written materials translated into the examinees' native

language and having translators knowledgeable in the necessary medical terminology present for oral simulations. It was important to eliminate any possible miscommunication or confusion due to language barriers when assessing the examinee's fund of knowledge and ability to manage complex case scenarios.

Another potential barrier was cultural. We attempted to create an environment that was professional, conducive to participation, and likely to be culturally acceptable to the examinees. Although the actual examinations were conducted by visiting physicians, we encouraged participation whenever possible by physician leadership from the host country. They gave input into the content covered and procedures used in the examination process. We also used information gathered by observing their actual clinical practice to guide creation of materials and scenarios that were as realistic as possible and sensitive to their cultural expectations. Moreover, after the first class of trainees was qualified, they were trained and used as examiners for the next class. Now that there are enough graduates who have been trained, the examiners are entirely Italian, which has made the administration of the examination easier, without need for translators, and more acceptable to the candidates.

Other potential barriers that needed to be addressed were the preconceived expectations of the learners and teachers involved. It was imperative to understand and appreciate the prior training experience and knowledge base of the physicians participating in the program. The learners in this case were highly trained adult learners, with a significant

Table 2 Rank preferences by type of examination: usefulness in assessing fund of knowledge and clinical ability (1 = most useful and 4 = least useful)

	Structured oral exam	Multiple-choice exam	Semi-structured oral exam	Essay exam	<i>p</i> value
Fund of knowledge ^a	1.55 ^b	2.68	2.77	3.00	$p<0.01$
Clinical ability ^c	1.61 ^b	2.73	2.57	3.00	$p<0.01$

^a Missing data: 2 of the 24 respondents did not respond to this question

^b Oral pre-test ranked significantly higher than the other three exams on both measures: fund of knowledge and clinical abilities. No other significant pairwise comparisons found

^c Missing data: 1 of the 24 respondents did not respond to this question

existing skill set that needed augmentation in specific areas (as they often are in similar projects). In this case they were well-trained in internal medicine and cardiology, but required further training in the acute care aspects of trauma surgery, orthopedics, wound care, pediatrics, ophthalmology, otolaryngology, obstetrics, and gynecology related to the practice of EM. In constructing the examinations for the program, we made every effort to address the areas that we felt needed to be covered in the greatest depth to ensure that the curriculum met the needs of the training program as defined by both the host and visiting country's leadership.

Limitations This study has several limitations. Learner perception is by definition subjective and an incomplete measure of an assessment tool. Learner satisfaction with structured oral examinations indicates only that these examinations were well-received, and we cannot make conclusions as to the efficacy or validity of these examinations in assessing the learner. Although there were several statistically significant findings, this study also had a small sample size and larger, more in-depth studies are needed to further investigate the topic. The examinations used in the curriculum were tailored to the specific needs of the program and therefore had not been previously validated. Another problem is that questions on the survey referring to oral versus written examinations sometimes did not distinguish between pre-test or post-test, which could have led to imprecise results. Finally, the external validity of this study depends in part on the applicability of our findings to other cultural environments. Learners who are not familiar with oral examinations may be less ready to accept novel (to them) methods of assessment. Since the learners in this study were primarily physicians working at an academic center, they may have been more receptive to these assessment methods than physicians working in a different setting would have been.

Conclusions

Physician learners participating in a cross-cultural, international training program in EM perceived all examination methods as useful, including structured oral case simulations, multiple-choice tests, semi-structured oral examinations, and essay tests. Learners ranked the structured oral case simulations as the most highly rated testing method and felt that oral examinations were better at assessing their clinical abilities when compared to written examinations. Oral case simulations can be useful assessment tools in an international medical training program. The results of this

study may be useful in guiding the development of training programs in countries with similar educational goals and clinical practice environments.

Conflicts of interest None.

References

1. Epstein RM (2007) Assessment in medical education. *N Engl J Med* 356(4):387–396
2. Townsend AH, McIlvenny S, Miller CJ et al (2001) The use of an objective structured clinical examination (OSCE) for formative and summative assessment in a general practice clinical attachment and its relationship to final medical school examination performance. *Med Educ* 35(9):841–846
3. Patil NG, Saing H, Wong J (2003) Role of OSCE in evaluation of practical skills. *Med Teach* 25(3):271–272
4. Jefferies A, Simmons B, Tabak D et al (2007) Using an objective structured clinical examination (OSCE) to assess multiple physician competencies in postgraduate training. *Med Teach* 29(2–3): 183–191
5. Govindan VK (2008) Enhancing communication skills using an OSCE and peer review. *Med Educ* 42(5):535–536
6. BLS Healthcare Provider Course (2008). Available via: <http://www.americanheart.org/presenter.jhtml?identifier=3011975>. Accessed 3 May 2008
7. ACLS Provider Course (2008). Available via: <http://www.americanheart.org/presenter.jhtml?identifier=3011972>. Accessed 3 May 2008
8. Pediatric Advanced Life Support Course – PALS (2008). Available via: <http://www.americanheart.org/presenter.jhtml?identifier=3012001>. Accessed 3 May 2008
9. Sauer J, Hodges B, Santhouse A et al (2005) The OSCE has landed: one small step for British psychiatry? *Acad Psychiatry* 29(3):310–315
10. Power DV, Harris IB, Swentko W et al (2006) Comparing rural-trained medical students with their peers: performance in a primary care OSCE. *Teach Learn Med* 18(3):196–202
11. Reinhart MA (1995) Advantages to using the oral examination. In: Mancall EL, Bashook PG (eds) *Assessing clinical reasoning: the oral examination and alternative methods*. American Board of Medical Specialties, Evanston, pp 31–39
12. Solomon DJ, Reinhart MA, Bridgham RG et al (1990) An assessment of an oral examination format for evaluating clinical competence in emergency medicine. *Acad Med* 65(9 Suppl):S43–S44
13. Bianchi L, Gallagher EJ, Korte R et al (2003) Interexaminer agreement on the American Board of Emergency Medicine oral certification examination. *Ann Emerg Med* 41(6):859–864
14. Wang N, Witt EA, Schnipke D (2006) Rejoinder: a further discussion of job analysis and use of KSAs in developing licensure and certification examinations. *Educational Measurement: Issues and Practice* 25(2)
15. Walubo A, Burch V, Parmar P et al (2003) A model for selecting assessment methods for evaluating medical students in African medical schools. *Acad Med* 78(9):899–906
16. Epstein RM, Dannefer EF, Nofziger AC et al (2004) Comprehensive assessment of professional competence: the Rochester experiment. *Teach Learn Med* 16(2):186–196

17. Norman GR, Van der Vleuten CP, De Graaff E (1991) Pitfalls in the pursuit of objectivity: issues of validity, efficiency and acceptability. *Med Educ* 25(2):119–126
18. Korhuis PT, Nekhlyudov L, Ziganshin AU et al (2002) Implementation of a cross-cultural evidence-based medicine curriculum. *Med Teach* 24(4):444–446
19. Kolb S, Reichert J, Hege I et al (2007) European dissemination of a web- and case-based learning system for occupational medicine: NetWoRM Europe. *Int Arch Occup Environ Health* 80(6):553–557
20. Wass V, Wakeford R, Neighbour R et al (2003) Achieving acceptable reliability in oral examinations: an analysis of the Royal College of General Practitioners membership examination's oral component. *Med Educ* 37(2):126–131
21. Ban KM, Pini R, Sanchez LD et al (2007) The Tuscan Emergency Medicine Initiative. *Ann Emerg Med* 50(6):726–732
22. American Board of Emergency Medicine (2008) Available via: <http://www.abem.org/public/>. Accessed 9 May 2008
23. Miller GE (1990) The assessment of clinical skills/competence/performance. *Acad Med* 65(9 Suppl):S63–S67
24. Kearney RA, Puchalski SA, Yang HY et al (2002) The inter-rater and intra-rater reliability of a new Canadian oral examination format in anesthesia is fair to good. *Can J Anaesth* 49(3):232–236
25. Davis MH, Karunathilake I (2005) The place of the oral examination in today's assessment systems. *Med Teach* 27(4):294–297
26. Swing SR (2002) Assessing the ACGME general competencies: general considerations and assessment methods. *Acad Emerg Med* 9(11):1278–1288
27. Anastakis DJ, Cohen R, Reznick RK (1991) The structured oral examination as a method for assessing surgical residents. *Am J Surg* 162(1):67–70
28. Papadakis MA (2004) The step 2 clinical-skills examination. *N Engl J Med* 350(17):1703–1705
29. Weiner SG, Kelly SP, Rosen P, Ban KM (2008) The eight Cs: a guide to success in an international emergency medicine educational collaboration. *Acad Emerg Med* 15:678–682

Sean Kelly is the Director of Graduate Medical Education at Beth Israel Deaconess Medical Center and an Assistant Professor at Harvard Medical School in Boston, MA. One of his major research interests is the effect of overcrowding on clinical teaching.

Scott Weiner is Director of Research at the Department of Emergency Medicine of Tufts Medical Center in Boston, MA. He is currently the chair of the American Academy of Emergency Medicine's International Committee. He served as a Regional Coordinator of the Tuscan Emergency Medicine Initiative.

Philip Anderson is an Assistant Professor of Medicine (Emergency Medicine) at Harvard Medical School. He is the Director of International Emergency Medicine at the Beth Israel Deaconess Medical Center Department of Emergency Medicine.

Julie Irish is the Director of the Office of Educational Research at Beth Israel Deaconess Medical Center and the evaluation specialist on the Donald Reynolds Foundation Grant for the Advancement of Geriatrics Education at Harvard Medical School.

Greg Ciottone is an Assistant Professor of Medicine and Chair of the Disaster Medicine Section at Harvard Medical School. He is currently the Director of the Division of Disaster Medicine at Beth Israel Deaconess Medical Center where he works clinically in the Department of Emergency Medicine.

Ricardo Pini is an Associate Professor of Medicine at the University of Florence, Italy. He is the coordinator of the "Harvard Emergency Medicine Project" and Director of the Intensive Care/Observation Unit in the Emergency Department at Azienda Ospedaliero-Universitaria Careggi in Florence, Italy.

Stefano Grifoni is the Director of the Emergency Medicine Unit at the Azienda Ospedaliero-Universitaria Careggi in Florence, Italy.

Peter Rosen is the founding editor of *Rosen's Emergency Medicine: Concepts and Clinical Practice* and the *Journal of Emergency Medicine*. He is a member of the Institute of Medicine, Senior Lecturer at Harvard Medical School, and Visiting Professor of Emergency Medicine at the University of Arizona School of Medicine.

Kevin Ban is an Assistant Professor of Medicine at Harvard Medical School and Director of the Tuscan Emergency Medicine Initiative at the Universities of Florence, Pisa, and Siena, Italy. He is also the director of a project to develop a trauma center at the Meyer Pediatric Hospital in Florence, Italy.