American Journal of Pharmaceutical Education **2005; 69 (3) Article 41.**

RESEARCH ARTICLES

Real-Time Standardized Participant Grading of an Objective Structured Clinical Examination

Cindy D. Stowe, PharmD, and Stephanie F. Gardner, PharmD, EdD

College of Pharmacy, University of Arkansas for Medical Sciences

Submitted September 10, 2004; accepted December 4, 2004; published May 10, 2005.

Objective. To determine the reliability of using standardized participant (SP) actors as real-time graders for an objective structured clinical examination (OSCE).

Methods. All students enrolled in *Therapeutics I* participated in a 3-station OSCE. Faculty and SP actors graded student performance in real time using a dichotomous performance checklist. Standard setting for the OSCE was done using the borderline method. Inter-rater reliability between the faculty members and SP was determined.

Results. Fifty-one students completed the morning examination and 30 completed the afternoon examination. The inter-rater reliability between the faculty members and SPs were 0.84, 0.88, and 0.93 for the morning examination and 0.90, 0.92, and 0.98 for the afternoon examination. The standard determined by faculty members and SPs for the morning examination was 66.9% and 63.8%, and for the afternoon examination was 72.6% and 73.1%, respectively.

Conclusion. These data support the use of SP actors as graders for comprehensive therapeutics OSCEs. The examination standard setting was inconsistent between the SP and faculty graders. Therefore, an alternative method of standard setting, such as the Angoff method, is required.

Keywords: inter-rater reliability, objective structured clinical examination, standardized participants, assessment

INTRODUCTION

The objective structured clinical examination (OSCE) is a method of testing that uses standardized participants (SPs) to measure the clinical competence of students and focuses on observable behaviors to determine outcomes. Standardized participants act as patients or other health care providers in a standardized encounter to assess student performance. Faculty members typically serve as graders for pharmacy-based OSCEs. Standardized participant actors can accurately and reliably assess student performance.[1-5] The National Board of Medical Examiners now requires an OSCE as part of medical licensure, utilizing SPs as graders.[1,4,5]

Although this method of assessment provides information difficult to obtain through traditional pencil-and-paper tests, it requires considerable financial resources and faculty time. A 3-case OSCE for approximately 80 students, with 10 minutes for preparation and 10 minutes for the encounter, takes approximately 7 hours to administer, with 9 examination rooms operating simultaneously. This involves a minimum of 9 real-time graders for the entire 7-hour testing period. By having SPs assess student performance, faculty graders would not be needed during the examination. This would make the examinations less costly and less time intensive for the faculty members. The objective of this study was to determine the reliability of using SPs as both actors and real-time graders for a therapeutics final examination using the OSCE format by comparing SP checklist scoring and examination standard setting to that of faculty members.

Corresponding Author: Cindy D. Stowe, PharmD. Address: 4301 W. Markham Street #522, Little Rock, AR 72205-7199. Tel: 501-364-1828. Fax: 501-296-1168. E-mail: stowecindyd@uams.edu

METHODS

Course Description

All students enrolled in PhPr 4625, *Therapeutics I*, in 2003 participated in this evaluation. *Therapeutics I* is a 5-hour required course in the spring semester of the second professional year of the curriculum that consists of both didactic and problem-based learning. *Therapeutics I* consists of modules that address the pharmacotherapy of renal, cardiovascular, and gastrointestinal disorders. The final examination for the course was a 5-case, 200-point examination, with 3 cases presented in the traditional OSCE format and 2 cases presented in written form. The therapeutic skills checklist for each of the 5 cases were worth 30 points each. The communication component was worth 50 points, with 25 points split between the 3 OSCE cases (8.3 points per case) and 25 points split between the 2 written documentation cases (12.5 points per case).

OSCE Case Development and Description

The application of the OSCE methodology at our institution has been previously described in great detail.[6,7] Case content was determined by carefully defining the specific practice competencies for each module of *Therapeutics I*. Cases were written by course instructors and the content was validated by external reviewers. The cases were written to assess the course objectives and specifically evaluated the students' abilities to evaluate laboratory findings, counsel a patient regarding nonpharmacologic and pharmacologic therapy, and make recommendations to another health care provider regarding pharmacotherapy, such as drug selection and monitoring. The cases were also designed to assess communication skills, including empathy and confidence.

Each case included directions to the student, directions to the SP, and performance criteria for evaluating student performance. The performance criteria were provided in the form of dichotomous checklists for both therapeutic and communication skills. The 5 to 8 therapeutic checklist items were individualized to each case, while the same 4 communication items were used for all cases. The communication items consisted of the following: "Student introduces self," "Student provides information with confidence," "Student is sensitive to the patient/situation," and "Student asks SP if he/she has any further questions." Standardized participants acting as patients were laypersons who had served as SPs for other OSCEs for the Colleges of Medicine, Pharmacy, Nursing, and/or health-related professions. Standardized participants acting as physicians were senior medical students. All SPs attended 2 training sessions prior to the examination. Case-naive pharmacy practice residents served as mock pharmacy students for SP training. The author of the case, faculty members familiar with SP training, and professional SP staff participated in the training of the SPs.

OSCE Examination Procedure

The Clinical Skills Center has 10 examination rooms, with audio and video access to all rooms from a central location in the facility. Nine examination rooms with 3 sets of cases conducted simultaneously accommodated groups of up to 18 students in 60-minute blocks of time that ran consecutively. Students were assigned to either the morning or afternoon examination. The examinations consisted of 3 SP cases in the areas of cardiovascular and gastrointestinal disorders. The content of the examinations was designed so that the topics and the SP types used for each OSCE weighted material appropriately. For each case, students were allowed 10 minutes for preparation and 10 minutes for the clinical encounter. College of Pharmacy faculty members and SPs graded student performance in real time in the Clinical Skills Center. All clinical encounters were videotaped.

Standard Setting Procedure

As previously described, the borderline procedure was used to determine the passing score for the morning and afternoon SP sections of the final examination.[7] For the borderline method, in addition to completing the individual therapeutic and communication checklist items, all graders provided an overall rating of "outstanding," "clear pass," "borderline," or "clear fail." The scores on all borderline cases were calculated by determining the percent of checklist items performed. The standard (ie, passing score) for each examination session (morning and afternoon) was the mean score of the borderline scores for that session. The standard was calculated for both the faculty members and the SPs scoring.

Statistical Analysis

Descriptive statistics were used to describe the faculty members and SP grading, SP error description, SP error rates, and standard setting for both the morning and afternoon examinations. Standardized participant errors of commission were those in which the SP gave credit for an item not performed, while SP errors of omission were those in which the SP did not give credit for an item performed. Inter-rater reliability between the faculty members and SP graders was measured for each case using the Spearman rank correlation coefficient, which evaluates the consistency of multiple raters. Inter-rater reliability may range from 0.0 (no agreement) to 1.0 (complete agreement).

RESULTS

Eighty-one students completed the OSCE examination for *Theraputics I*, with 51 taking the morning examination and 30 taking the afternoon examination. The morning examination covered the topics of atrial fibrillation (AF), hepatic encephalopathy (HE), and stress ulcer prophylaxis (SUP). The afternoon examination consisted of cases involving stable angina (SA), ulcerative colitis (UC), and peptic ulcer disease (PUD). Four of the cases involved encounters with a physician SP (HE, SUP, UC, and PUD) and 2 cases involved encounters with a patient SP (AF and SA). Standardized participants had adequate time to complete all checklist items for all students.

There were 2,541 checklist items with 119 discrepancies observed between the SP and faculty checklist

Table 1. SP Checklist Item Discrepancy

	Therapeutic	Communication	Total
Number of items	1569	972	2541
Number of errors (%)	65 (4.1)	54 (5.6)	119 (4.7)
Number of Errors of Commission (%)	46 (70.8)	29 (53.7)	75 (63.0)
Number of Errors of Omission	19 (29.2)	25 (46.3)	44 (37.0)

Table 2. Summary of Checklist Scoring Results and Inter-rater Reliability

Therapeutic Content	Items	SP*Mean (SD)	Faculty*Mean (SD)	Inter-rater reliability†
Morning examination (n=51)				
Atrial fibrillation	10	28 (6)	27 (7)	0.88
Hepatic encephalopathy	9	32 (6)	31 (7)	0.84
Stress ulcer prophylaxis	12	33 (5)	33 (5)	0.93
Afternoon examination (n=30)				
Stable angina	11	35 (6)	34 (6)	0.90
Ulcerative colitis	12	34 (4)	34 (4)	0.92
Peptic ulcer disease	9	35 (5)	35 (5)	0.98

*Maximum points = 38.3

†Spearman rank correlation coefficient

items (Table 1). This resulted in an SP scoring accuracy rate of 95.3%. The distribution of SP discrepancies between the therapeutic and communication portion of the checklist was similar. However, the majority of discrepancies (63.0%) were in favor of the student, ie, an error of commission rather than omission.

The overall case performance scores were similar between the faculty and SP graders (Table 2). The average performance for the class as scored by the faculty and SP graders for the morning examination was 78.9% and 80.8%, respectively, and for the afternoon examination, 89.1% and 90.1%, respectively, which is consistent with the greater rate of errors of commission by SPs. Good to excellent agreement between faculty and SP graders was observed (0.84-0.98, Table 2).

The borderline method for standard setting determined by faculty members and SPs on the morning examination was 66.9% and 63.8%, respectively. The afternoon examination standard determined by faculty members and SPs was 72.6% and 73.1%, respectively. Faculty and SP graders identified a similar number of encounters as borderline performances. However, the faculty and SP graders differed on the encounters rated as "borderline." The difference in the standard between faculty and SP graders for the morning examination was felt to be unacceptable.

DISCUSSION

This paper outlines the use of SPs as real-time graders for a pharmaceutical-based OSCE. Standardized participant graders proved to be capable of scoring student performances at an acceptable level of agreement with faculty graders for an OSCE. Standardized participants have demonstrated accuracy and reliability in grading medical student OSCEs using traditional checklists.[1-5] These data are consistent with the findings of Heine et al who found that when SPs make checklist scoring errors, they tend to do so in favor of the student (Table 1).[2] Standardized participants may actually be better at the assessment of communication within these encounters than faculty members.[8,9] The perspective of participant versus observer as well as patient versus teacher may lead to a more accurate assessment of communication. This may be why the discrepancies in the SP checklist items for communication were equally split between errors of commission and omission (Table 1). However, SPs performed less well at the borderline method of standard setting. This method relies on a global assessment of performance, and under the conditions outlined here, SPs were unable to set a standard for the examination that was consistent with that of the faculty members.

Although the application of the OSCE methodology using a 10-minute preparation and 10-minute clinical encounter time for pharmacy students is different from the application employed by the US Medical Licensing Examination (USMLE) Step 2 Clinical Skills (CS) examination, the fundamental use of SPs is the same. The 2004 Step 2 CS examination uses cases that may be staged as face-to-face or telephone encounters with patients or caregivers.[5] For each case, the encounter, including review of the examination instruction sheet, comprises up to 15 minutes, followed by at least 10 minutes for written documentation of the encounter. The SP scores the student on

the following: history taking, physical examination, and communication/interpersonal skills. Standardized participants complete the encounter checklist at the conclusion of each encounter. Data support the validity and reliability of such an examination.[1,4,5]

While there are adequate data to support the skills of an SP scoring checklist at the end of the clinical encounters,[1,2,4,5] the use of SPs to provide a global assessment of the encounter has not been established.[10] In fact, global rating scores given by expert graders more accurately differentiated levels of mastery.[10,11] The most compelling limitation for SPs in this aspect of assessment is their lack of expertise beyond the therapeutic checklist items. The borderline method depends on a global assessment of the encounter to identify the minimally acceptable performances.[7] The agreement between global ratings is as high as the agreement between checklist items if the grader making the global assessment has adequate expertise.[11-13] The SP graders in these cases were senior medical students and laypersons without medical training. If the graders lack a sufficient level of expertise then the checklist items could be weighted to allow a more accurate global assessment.[11]

Therefore, to overcome the lack of experience and expertise of the SPs, an alternative method of standard setting should be employed. The Angoff and borderline method of standard setting for pharmacy OSCEs have been demonstrated to be equivalent.[7] From a resource management point of view, the use of a weighted checklist may potentially offer the most advantage given that the SPs are grading in real time without faculty members present. The use of the Angoff method requires more faculty member time, albeit less than is required for faculty members to provide real-time grading of the OSCE.

An additional concern to consider with the expanded responsibilities of SPs as actors and graders will be the level of fatigue that can exist during long periods of testing and the provision of adequate time for SP checklist scoring. Therefore, we have limited our SPs to either the morning or afternoon examination, built in at least 2-3 breaks during each examination, and allowed an additional minute between the student encounters to give the SP additional time to review the checklist items. Paperless grading using tablet personal computers and *WebSP* software (Lionis, Inc, Hungary) further ensures the ability of the SPs to complete the necessary tasks required in the encounter. Other factors such as checklist length, checklist clarity and appropriate reading level, and intensive SP training that utilizes "practice students" have contributed to increased SP accuracy in checklist scoring. Further study is needed to determine whether SPs' assessment of the global performance of pharmacy students can be enhanced through checklist weighting and more training focused at identifying borderline performances. As was expected, SPs were capable of achieving excellent agreement with faculty graders on both therapeutic and communication skill checklists.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the valuable input of faculty case authors and reviewers of the OSCEs because without their input and support this work could not happen. In addition, OSCEs are possible because of the University of Arkansas for Medical Sciences Clinical Skills Center and its professional staff.

REFERENCES

1. Adamo G. Simulated and standardized patients in OSCEs: achievements and challenges 1992-2003. *Med Teach*. 2003;25:262-70.
2. Heine N, Garman K, Wallace P, Bartos R, Richards A. An analysis of standardised patient checklist errors and their effect on student scores. *Med Educ*. 2003;37:99-104.
3. Pangaro LN, Worth-Dickerson H, Macmillan MK, Klass DJ, Shatzer JH. Performance of "standardized examinees" in a standardized-patient examination of clinical skills. *Acad Med*. 1997;72:1008-11.
4. An analysis of US student field trial and International Medical Graduate Certification Testing Results for the proposed USMLE clinical skills examination. Available at: http://www.usmle.org/news/Step2CSNews/Step2ftresults2503.htm. Accessed September 3, 2004.
5. 2004 USMLE Step 2 Clinical Skills Content description and general information. A Joint Program of the Federation of State Medical Boards of the United States, Inc., and the National Board of Medical Examiners. Available at: http://www.usmle.org/step2/Step2CS/Step2CS2005GI/toc.asp. Accessed September 3, 2004.
6. Monaghan MS, Vanderbush RE, Gardner SF, Schneider EF, Grady AR, McKay AB. Standardized patients: an ability-based outcomes assessment for the evaluation of clinical skills in traditional and nontraditional education. *Am J Pharm Educ*. 1997;61:337-44.
7. Gardner SF, Stowe CD, Hopkins DD. Comparison of traditional testing methods and standardized patient examination for Therapeutics. *Am J Pharm Educ*. 2001;65:236-40.
8. Cooper C, Mira M. Who should assess medical students' communication skills: their academic teachers or their patients? *Med Educ*. 1998;32:419-21.
9. Sibbald D. Using first-year students as standardized patients for an objective structured clinical exam for third-year pharmacy students. *Am J Pharm Educ*. 2001;65:404-12.
10. Hodges B, McIlroy JH. Analytic global OSCE ratings are sensitive to level of training. *Med Educ*. 2003;37:1012-6
11. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med*. 1998;73:993-7.
12. Cunnington JPW, Neville AJ, Norman GR. The risks of thoroughness: reliability and validity of global ratings and checklists in an OSCE. *Adv Health Sci Educ*. 1997;1:227-33.
13. Wilkinson TJ, Frampton CM, Thompson-Fawcett M, Egan T. Objectivity in objective structured clinical examinations: checklist are no substitute for examiner commitment. *Acad Med*. 2003;78:219-23.