

Comparison of Traditional Testing Methods and Standardized Patient Examinations for Therapeutics

Stephanie F. Gardner, Cindy D. Stowe and Denise D. Hopkins

College of Pharmacy, University of Arkansas for Medical Sciences, Slot 522, 4301 West Markham Street, Little Rock AR 72205-7122

Standardized patient (SP) examinations require that a student demonstrate problem solving, comprehension, and communication skills. The primary objective of this study was to determine if there is a correlation between performance on SP examinations and traditional testing methods in a therapeutics course. A secondary objective was to compare three standard-setting procedures for an SP examination. A final examination, consisting of a 75-item multiple choice examination and a three-station SP final examination, was administered to 73 pharmacy students. Three standard setting procedures (Angoff, borderline, and holistic) were applied to the SP examination data. A moderate positive correlation was seen between performance on the multiple-choice and SP final examinations. There was a weak correlation between performance on interim examinations in the three content domains and performance on the individual SP cases related to that domain. The Angoff and borderline procedures gave similar results, though the borderline method provides a less labor-intensive approach. An SP final examination provided additional information regarding students' problem-solving, clinical reasoning, and communication skills at the culmination of therapeutics.

INTRODUCTION

Students in most medical schools now undergo standardized patient (SP) based assessments(1). The United States Medical Licensing Examination has done extensive research to determine the feasibility of incorporating SP examinations into their licensing procedures for physicians, in a manner similar to the licensing procedure for physicians in Canada. SP examinations are now a compulsory part of the licensing examination for Canadian pharmacists, through the Pharmacy Examining Board of Canada (PEBC)(2). These examinations are also used as a part of continuing competency testing for practicing pharmacists in the province of Ontario(3). Although this form of assessment is not routinely used for licensing or continuing competency assessment for pharmacists in the United States, the emphasis on competency-based curricula and evaluation by the American Council on Pharmaceutical Education (ACPE) stresses the need for continued research in this area. ACPE recommends clinical evaluation measure "cognitive learning, mastery of essential practice skills, abilities to communicate effectively and to use data in the critical thinking and problem solving processes"(4). The objective structured clinical examination (OSCE) is a method that utilizes SPs to measure clinical competence by focusing on outcomes through observable behaviors.

Although preparing students to become health care providers who are capable of providing pharmaceutical care in a variety of practice settings is the primary mission of pharmacy education, the consistent measurement of a student's ability to provide this type of care remains elusive. The problem solving skills and communication skills that must accompany a sound knowledge base are difficult to assess. Recently, the American Association of Colleges of Pharmacy (AACP) pro-

posed that the Pharmacy College Admissions Test (PCAT) include an essay component to evaluate critical thinking skills(5). The AACP PCAT Advisory committee recommended an essay component, rather than a multiple-choice format, secondary to its ability to not only assess critical-thinking, but also writing skills. The perceived need to assess critical-thinking and communication, even prior to admission, is evidence of the shift toward ability-based outcomes measurement in pharmaceutical education.

The objectives of this study were to compare the scores on individual short-answer interim examinations and corresponding cases on the standardized participant portion of the final examination, as well as to compare the scores from two parts of a final examination: a 75-item multiple-choice portion and a three-case standardized patient based portion. A secondary objective of the study was to compare three standard-setting procedures for a standardized patient examination for pharmacy students.

METHODS

Student Selection

All students enrolled in Therapeutics I participated in this evaluation. Therapeutics I is a five-hour required course in the spring semester of the second professional year of the curriculum. The focus of Therapeutics I is the use of drug therapy in treating major disease states, including a review of pathophysiology of the disease, therapeutic goals, plans of treatment, dosage regimens, therapeutic alternatives, and therapeutic endpoints. The course consists of four 50-minute lecture hours weekly, in addition to a weekly recitation laboratory in which

Am. J. Pharm. Educ., 65, 236-240(2001); received 6/19/01, accepted 8/17/01.

the class is divided for small group discussion and case presentation. A total of 74 students were enrolled in the spring semester of 2001.

MC Examination

A 75-item multiple-choice computerized examination was administered in a two-hour block to students in Therapeutics I. This examination was scheduled three days after the SP examination during finals week.

SP Examination

Station Development. Station content was determined by carefully defining the specific practice competencies of each module of Therapeutics I. Modules included fluid and electrolyte, renal, cardiovascular, and gastrointestinal disorders. Stations were developed with the aid of instructors for each of these modules and the staff of the standardized participant program on our campus. The authors of the stations (module instructors) developed cases based on their practice experience that focused on the primary competencies covered in their modules. The cases were content-validated by an expert at three external academic institutions. The three individuals chosen for content validation were clinical pharmacy practitioners with experience in assessing student performance through traditional methods and SP testing methods. One individual was a faculty member at another college of pharmacy and the other two were employed by hospitals, but worked in ambulatory clinic settings. The content validation panel was asked to apply three practical questions that all instrument items should meet: (i) Can each checklist item be performed exactly as it is written? (ii) Will the checklist items mean the same thing to each evaluator? (iii) Is each checklist item relevant and necessary for the case?

Each station included directions to the student, directions to the SP, and performance criteria for evaluating the student. A sample station is shown in Appendix A. Because the pharmacist is often involved in providing expertise to other health care providers, the SP in pharmaceutical education is more often a "standardized participant"⁽⁶⁾. The standardized participant may be a physician, nurse, or patient with whom the pharmacist interacts. SPs were trained prior to testing to ensure a standardized response to all students. For stations involving an encounter with a patient, actors portrayed the patients. For stations involving a physician, third and fourth-year medical students served as SPs. The performance criteria were provided in the form of checklists. The checklists contained behaviorally anchored items that assessed application of therapeutic knowledge and basic communication skills. Faculty members graded student performance in real-time in the Clinical Skills Center and videotape back-up provided the potential for review after initial grading in the event that a grade was contested by a student.

Examination Procedure. Students were randomly assigned to the morning or afternoon examinations, which consisted of three SP cases in the areas of renal, cardiovascular, and gastrointestinal disorders. The morning cases covered the topics of diabetic nephropathy, hypertension, and peptic ulcer disease. The afternoon cases consisted of renal dosage adjustment, heart failure, and stress ulcer prophylaxis. For each SP case, students were allowed 15 minutes for preparation and 15 minutes for the clinical encounter.

Fifteen graders, nine for the morning session (approx-

mately five hours) and six for the afternoon session (approximately 3 1/2 hours), assisted with assessment of the SP examination. Of the 15 graders, there were 11 faculty members, two residents, and two fellows. Ten of the 15 graders had previous experience grading SP examinations. The five graders without previous experience were oriented to the grading procedure by reviewing checklists with the station authors. Station authors were available during testing for questions.

Standard Setting Procedure

The 75-item multiple-choice final examination followed the holistic procedure for standard setting. The policy of the University of Arkansas for Medical Sciences College of Pharmacy Department of Pharmacy Practice states that a passing score is 70 percent. As previously described in the medical literature, the Angoff, borderline, and holistic procedures were used to determine the passing score for the morning and afternoon SP sections of the final examination⁽⁶⁾. Each of the three cases was equally weighted as 25 percent of the grade and communication skills assessed during each case were pooled and weighted as 25 percent.

For the Angoff procedure, five faculty members reached consensus on a definition of a minimally acceptable "borderline" candidate. With this definition in mind, each faculty member rated each checklist item independently, answering the following question: "What proportion of borderline candidates will be able to successfully pass this checklist item?" After individual rating, the group discussed the checklist items and their ratings and explained their reasoning when discrepancies occurred. Following this discussion, all faculty members, answering the same question performed a second rating. The average of the second set of scores was used to determine the passing scores for the morning and afternoon SP sections of the final examination.

For the borderline method, each faculty grader, in addition to completing the station checklist, provided an overall rating of "outstanding," "clear pass," "borderline," or "clear fail"⁽⁷⁾. All cases that received a rating of "borderline" were reviewed. A percentile score was calculated for each borderline case. The passing scores for the morning and afternoon sessions were based on the mean score of the identified borderline cases within those sessions.

Statistical Analysis

Descriptive statistics were used to represent the data. Correlation analysis was done to determine if a linear association existed between the traditional testing methods of short-answer and multiple choice to the objective structured clinical SP examination in total and by station. Nonlinear relationships between variables were assessed by visualization of the scatterplots to determine if there was evidence of a curve. The level of significance was set at an $\mu = 0.05$. Data was represented as mean (\pm SD).

RESULTS

Students

Fifty students completed the SP examination in the morning and 23 students completed the afternoon session. One student was unable to take the examination secondary to medical reasons.

Standard Setting

The Angoff procedure yielded a mean (\pm SD) passing

Table I. Standard setting results and student performance

Examinations	Percent passing score		Number of failures		Percent failure rate	
	AM	PM	AM	PM	AM	PM
SP						
Angoff	61	67	13	6	26.0	26.1
Borderline	62	67	15	6	30.0	26.1
Holistic	70	70	27	8	54.0	34.8
Multiple-Choice	70		29		39.7	

score of 61 percent (± 20.8) for the morning examination session and a mean passing score of 67 percent (± 17.6) for the afternoon session. Performance on the SP examination was scaled and reported to the students based on the Angoff procedure. Student performance is shown in Table I. Thirty-seven of the 150 case encounters (24.7 percent) in the morning session were rated as borderline. Seventeen of the 69 case encounters (24.6 percent) in the afternoon session were rated as borderline. The borderline method yielded a mean (\pm SD) passing score of 62 percent (± 8.8) for the morning session and a mean passing score of 67 percent (± 13.1) for the afternoon session. Using the holistic method, a passing score of 70 percent was applied to the SP examination, resulting in a failure rate of 47.0 percent (35/73 students).

Data Analysis

Performance on interim examinations was compared to performance on individual station cases of the SP examination. The mean (\pm SD) communication score on the SP examination was 90.2 (± 10.7). Only four students scored below 70 percent on this portion of the SP examination. There was no statistically significant positive correlation between interim examination scores and performance on individual SP station cases.

Using the Angoff standard setting procedure to scale the SP examination grades, the mean score for the morning session was 77.4 percent (± 13.5), as compared to a mean score of 75.0 percent (± 14.6) for the afternoon session. Using the borderline method, similar results were obtained with a mean score for the morning session of 76.2 percent (± 13.3) and a mean score of the afternoon session of 75.0 percent (± 14.6).

After scaling the SP examination with the Angoff method, the mean (\pm SD) scores on the final examinations were 76.2 percent (± 13.8) and 72.2 percent (± 11.4) for the SP and multiple choice portions, respectively. The Pearson correlation coefficient (r) for the relationship between the performance on the two examinations was 0.58 ($P < 0.0001$). The distribution of letter grades on the two examinations is shown in Figure 1. Although the averages on the two examinations were similar, the distribution of letter grades was more normally distributed for the SP examination.

DISCUSSION

Traditionally, colleges of pharmacy have assessed students using objective, short-answer, and multiple-choice examinations. However, clinical faculty members often see a disparity between high classroom achievers and achievers in the clinical setting. This disparity may stem from the process of testing memorization versus clinical application of knowledge. SP examinations have been identified as useful teaching and evaluation tools during pre-clinical coursework in medical schools(8). A commentary attesting to their use stated that "The purpose of performance assessment is to provide behavior-based evidence that skills have been mastered. In contrast,

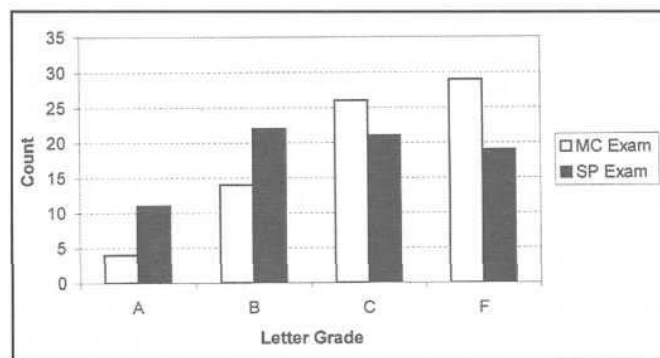


Fig. 1 Letter grade distribution for multiple-choice (MC) and SP final examinations in Therapeutics I.

knowledge-based assessment (such as multiple-choice exams) allows broad sampling of a student's cognitive database, while providing only limited insight on ability to apply knowledge in a clinical setting."(9) In an effort to better assess clinical reasoning and problem solving skills, an SP examination was added to our therapeutics course.

We compared various standard-setting procedures for SP examinations in this study. We agreed, at the outset, to use the Angoff procedure for assignment of grades to the SP portion of the final examination. This decision was made because we were not sure that we would have enough cases identified as "borderline" given the limited number of stations. We could not find evidence in the literature that indicated the holistic method was valid for SP-based assessment. The results of this study showed that the borderline and Angoff standard-setting procedures gave very similar results. The failure rate seen with these procedures (*i.e.*, 26-30 percent) is similar to historical final examination performance using our traditional testing procedures. The holistic method resulted in an unacceptable failure rate of 47 percent (Table I). The Angoff procedure required six faculty members for a three-hour discussion, followed by data input and analysis, to determine passing score. In contrast, the borderline method was done concurrently with grading of the SP examination and required little additional effort to determine the passing score.

The results of this investigation showed that there was a moderate positive correlation between our traditional multiple-choice format final examination scores in a therapeutics course and a SP final examination. We would expect that a correlation would be seen, as therapeutics knowledge is a major component of both testing methods. However, the added assessment of critical thinking and communication skills on an SP examination may better identify students who will excel in the clinical environment, secondary to their ability to apply the knowledge that they have obtained.

Performance on individual station cases within the SP

examination was only marginally correlated with performance on interim examinations in that content area. This finding may be secondary to the fact that each case was written to assess a small number of competencies within the area. A larger number of cases within each module may have strengthened the relationship between performance on interim examinations and performance on cases in that domain.

Although the primary focus of the SP examination was the assessment of clinical reasoning, communication skills were directly and indirectly assessed. Four students failed to obtain a 70 percent on the communication checklist items. Direct assessment of communication skills came from checklist items directed at the students' level of confidence and respect for the SP. This information could be used to provide formative assessment to the individual students prior to progression. Indirect assessment of communication skills was achieved by assessing students' ability to cover essential information in a given period of time. It was also achieved by assessing the students' ability to adjust the complexity and sophistication of their responses to the situation at hand.

This study demonstrated that a reasonably fair and defensible standard pass rate can be determined for SP examinations in therapeutics using the Angoff or borderline methods. Although a larger number of stations has been shown to increase the reliability of OSCEs in a high stakes environment (e.g., national licensing examination), a three-station SP examination provided us with an economically feasible way to better assess critical thinking and communication skills. Verhoeven and colleagues demonstrated that shortening the SP examination by using fewer stations could be achieved without a loss in reliability by the addition of a separate computer-administered true/false examination(10). In order to use SP examinations exclusively for assessment, a larger number of stations would be required to ensure reliability. However, combining SP examinations with objective, computerized multiple-choice examinations, provides a more global assessment of student performance in a therapeutics course.

Acknowledgements. The authors gratefully acknowledge the assistance of Nancy Carthan, PharmD, Elizabeth Miller, PharmD, and Michael Monaghan, PharmD, for their expertise in reviewing the SP cases.

References

- (1) Stillman, P.L., Regan, M.B., Philbin, M. and Haley, H.L., "Results of a survey on the use of standardized patients to teach and evaluate clinical skills," *Acad. Med.*, **65**, 288-292(1990).
- (2) The Pharmacy Examining Board of Canada. Available online at <http://www.pebc.ca/>.
- (3) Fielding, D.W., Page, G.G., Rogers, W.T., O'Byrne, C.C., Schulzer, M., Moody, K.G., "Standard setting for a test of pharmacy practice knowledge: Application in high-stakes testing," *Am. J. Pharm. Educ.*, **60**, 20-29(1996).
- (4) American Council on Pharmaceutical Education, "Accreditation standards and guidelines for the Professional Program in Pharmacy Leading to the Doctor of Pharmacy Degree," June 14, 1997.
- (5) Zoeller, J.L. (edit.), "AACP proposes added critical thinking component to PCAT," *AACP News*, **32**(5), 3(2001).
- (6) Monaghan, M.S., Vanderbush, R.E., Gardner, S.F., Schneider, E.F. and Grady, A.R., "Standardized patients: An ability-based outcomes assessment for the evaluation of clinical skills in traditional and nontraditional education," *Am. J. Pharm. Educ.*, **61**, 337-344(1997).
- (7) Kaufman, D.M., Mann, K.V., Muijtjens, A.M.M. and der Vleuten, C.P.M., "A comparison of standard-setting procedures for an OSCE in undergraduate medical education," *Acad. Med.*, **75**, 267-271(2000).
- (8) Barrows, H.S., "An overview of the uses of standardized patients for teaching and evaluating clinical skills," *ibid.*, **68**, 443-451(1993).
- (9) Anonymous, "Evaluating the internal medicine clerkship: a CDIM com-

mentary. Clerkship Directors in Internal Medicine," *Am. J. Med.*, **97**, I-VI, (1994).

- (10) Verhoeven, B.H., Hamers, J.G.H.C., Scherpbier, A.J.J.A., Hoogenboom, R.J.I. and van der Vleuten, C.P.M., "The effect on reliability of adding a separate written assessment component to an objective structured clinical examination," *Med. Educ.*, **34**,525-529(2000).

APPENDIX A.

Station Name: Renal Dosage Adjustment

Author: Michael A. Marx, PharmD

Task: Recommend dosage adjustment to prescriber based upon creatinine clearance determination.

Station Requirements: APhA Drug Information Handbook

Instructions for the Student

A physician approaches you with questions about adjusting medication doses based on a patient's renal function.

Patient Information:

The patient is a 5'6" female with chronic hepatitis B that was admitted to the ICU one week ago following an automobile accident. She is recovering satisfactorily from her traumatic injuries but during surgery to repair a ruptured spleen, she became profoundly hypotensive. This persisted for about 24 hours. Urine output was initially decreased but is now sufficient to maintain fluid balance.

Her serum creatinine was 1.0 mg/dL upon admission and steadily increased to 2.5 mg/dL over 4 days where it has remained. Her most recent weight is 60 kg.

Her medications on admission were:

Lamivudine 100 mg QD
Multivitamin QD
Albuterol Inhaler 2 puffs PRN

The following medication has been added since admission

Ranitidine 50 mg IV q8h
Piperacillin 4 gm IV q6h

Urine Data for past 24 hrs

Urine volume: 1800 ml
Urine creatinine: 66 mg/dL

Instructions for the Standardized Patient

You are Dr. Brown, a new intern that has just been assigned to this patient. Upon review of her information, you realize that the patient has had some degree of kidney problems since she underwent surgery to remove a ruptured spleen. You have not yet met the patient and you know very little else about her.

The student should provide you with any necessary dosage adjustments of the medications that she is receiving.

- If the student does not provide this information please ask the student the following: "Do I need to make any dosage adjustments of her current medications?"
- If the student does provide this information, they may or may not volunteer how they came up with these changes. If the student provides you with the creatinine clearance, acknowledge that you thought it was decreased. If the student doesn't provide you with the creatinine clearance information, please ask the student "What did you base your dosage recommendations on?" (CICr = 33-34 ml/min/1.73m²)

If asked, you may offer the following information:

1. BB has no known drug allergies.
2. You do not know BB's age.
3. Her infection is 'serious'.

Checklist

	Fail	Pass		O	O
Student recommends adjusting the dose of lamivudine.	O	O	any where from 50 IV Q12h to 50 IV 24h)	O	O
Student recommends the correct renal adjusted dose of lamivudine (50 mg QD)	O	O	Student recommends adjusting the dose of piperacillin	O	O
Student recommends no dosage adjustment for the multivitamin.	O	O	Student recommends the correct renal adjusted dose of piperacillin (3-4 gm q8hrs)	O	O
Student recommends no dosage adjustment for the albuterol.	O	O	Student states the calculated creatinine clearance (CICr = 33 ml/min or 34 ml/min/1.73m2)	O	O
Student recommends adjusting the dose of ranitidine	O	O	Student introduces self 41	O	O
Student recommends the correct renal adjusted dose of ranitidine (50-100 mg over 24 hours this maybe	O	O	Student provides information with confidence	O	O
			Student is sensitive to needs of patient/situation	O	O
			Student asks the standardized participant if she/he has any further questions.	O	O

