REGULAR ARTICLES

Development of a Comprehensive Examination for Baccalaureate Pharmacy Students Prior to Experiential Training

Maureen Smythe
Richard L. Slaughter
Mary Jane Sudekum

INTRODUCTION

Traditionally, the B.S. curriculum within colleges of pharmacy is designed to provide students with two years of didactic instruction prior to experiential training. The underlying assumption is that students need basic knowledge in the areas of medicinal chemistry, pharmacology, pharmacokinetics, and therapeutics to function successfully in an experiential environment. Preceptors often assume

Maureen Smythe, Pharm.D., is Assistant Professor (Clinical), Department of Pharmacy Practice, 328 Shapero Hall, Wayne State University, Detroit, MI 48202, and Clinical Education Coordinator, Department of Pharmaceutical Services, William Beaumont Hospital, Royal Oak, MI. Richard L. Slaughter, M.S., is Professor and Chairman, Department of Pharmacy Practice, Wayne State University. Mary Jane Sudekum, Pharm.D., is Assistant Professor (Clinical), Department of Pharmacy Practice, Wayne State University, and Clinical Pharmacist, Department of Pharmacy Services, Harper Hospital, Detroit, MI.

53

that students who perform poorly on rotations have struggled in earlier course work. Many colleges of pharmacy require that students take a comprehensive examination prior to entering into experiential training. Several investigators have reported that performance on preclinical exams is not reflective of performance during clinical rotations (1-3). Gehres and colleagues, however, concluded that a comprehensive examination prior to experiential training may identify students who could benefit from remediation (3).

A comprehensive examination has been administered to fifth-year pharmacy students at Wayne State University prior to experiential training for the past six years. The purpose of this examination has been to evaluate individual areas of student weakness that may be strengthened on rotations. The questions contained in this examination have been submitted by the faculty. However, the content has not undergone internal or external review, and the exam has not been proven valid or reliable. In addition, preceptors have only been provided with an overall score, which provides no means of identifying specific areas of weakness.

The objective of this project was to develop a valid, reliable comprehensive examination that could be used to identify individual areas of student weakness prior to experiential training.

METHODOLOGY

Examination Development

The first step in developing the comprehensive examination was to request questions for possible inclusion from the faculty of pharmacy. The investigators solicited multiple-choice questions written with five distracters from individual faculty/adjunct faculty members in their areas of expertise. Nine faculty members from the Department of Pharmacy Practice submitted questions from ten therapeutic areas, including congestive heart failure (CHF), hypertension (HTN), arrhythmias, seizures, urinary tract infections (UTIs), pneumonia, diabetes, asthma/chronic obstructive pulmonary disease (COPD), and thromboembolic disease/anemia. Six faculty members from the Department of Pharmaceutical Sciences submitted ques-

tions in the areas of pharmacokinetics, pharmacology, and medicinal chemistry. Faculty submitted a total of 324 questions. The content of questions submitted was consistent with the content of questions that typically appear on examinations within our college. Exam questions were aimed at testing ability both to apply concepts and to recall significant information.

An internal panel of reviewers consisting of pharmacy faculty members reviewed all examination questions for content validity. Content validity is the degree to which the test includes a representative sample of all tasks that could have been included. Each examination question was reviewed independently by two faculty members within the same discipline, except for three circumstances when faculty members from different disciplines shared an expertise. Faculty members were not permitted to review their own questions. Content validity was measured using the method of Pray and Popovich (4). Questions were ranked for content validity on a scale of one to three as follows:

1. The question is very representative of the knowledge or skills required in the indicated area.
2. The question is possibly representative of the knowledge or skills required in the indicated area.
3. The question is not representative of the knowledge or skills required in the indicated area.

Questions with content validity rankings of one or two or a combination thereof were included in the initial testing tool. Exam questions with a ranking of three by both reviewers or one ranking of two and one ranking of three (a total of 11 questions) were excluded from inclusion in the initial examination. A second internal panel of reviewers consisting of two faculty members (also part of the first review panel) reevaluated all questions with a content validity ranking of three and one (a total of 22 questions). The majority of these questions were either excluded or revised as deemed necessary by the second review panel. Upon evaluation by the second review panel, eight questions with an original ranking of three and one were included in the initial testing instrument as originally written by the faculty member. A total of 300 questions

with content validity rankings of 1's or 2's either initially or upon revision were used to develop 2 separate 150-point examinations (Exam A and Exam B). Questions were split into the two exams based upon topic. Each examination included 14 pharmacology questions, 13 medicinal chemistry questions, 13 pharmacokinetics questions, and 11 questions from each of the 10 previously mentioned therapeutic topics. Each examination question was assigned a point value of one.

Field Testing

The next step in developing the examination was to seek out pharmacy students across the country who would be interested in participating in a field test of the initial examination. Letters to department chairs or deans were mailed out to all colleges of pharmacy in the United States that offer a traditional 5-year B.S. degree in pharmacy, a total of 54 colleges. A total of 16 colleges of pharmacy responded to the letter, for a 30% response rate. Of these 16 schools, 9 indicated that they would be able to recruit students to participate in field testing the examination. Examinations, along with a test administrator's guide and specialized scoring forms, were mailed to each college. Students were allowed two hours to complete the exam. Test administrators were instructed to account for all exams and scoring forms. Exam scoring forms were returned by mail, and all exams were scored by the university testing service.

The results of the field test were used to perform the individual item analysis to identify questions of the appropriate difficulty and discrimination for inclusion in the final test instrument. Item difficulty is the percentage of exam takers who chose the correct response. This statistic follows an inverse relationship: the higher the difficulty, the easier the test item. For 5-option multiple-choice questions, optimal difficulty is 60%, while most of the items should fall into the 50-80% difficulty range. Questions within this range have the greatest potential to differentiate among students with different levels of knowledge. A measure of item discrimination, the point biserial correlation coefficient, was assessed to distinguish between those students who know the material and those who are guessing. Ideally, the correct response should have a positive point biserial

correlation. Positive correlations indicate stronger students are selecting the correct answer more often than weaker students. Only questions that fell into the 30-80% difficulty range with a positive point biserial correlation coefficient ≥ 0.2 for the correct response were considered for inclusion in the final test instrument.

RESULTS

Each exam was field tested on pharmacy students from five different colleges of pharmacy (one college field tested both exams). Students participating in the field test were near completion of their fourth year of a traditional five-year B.S. program. Exam A and Exam B were field tested on 174 and 140 students, respectively. A total of 148 questions were scored on Exam A. Two questions were dropped prior to scoring, one for a typographical error and the other for an incorrect structure. A total of 149 questions were scored on Exam B; 1 question was dropped because of a typographical error. Descriptive statistics summarizing performance on both exams are listed in Table 1. There was no significant difference in overall examination scores between Exam A and Exam B; however, there were significant differences between Exam A and Exam B in subtest scores in the areas of diabetes and

TABLE 1

Statistic	Exam A	Exam B
Number of questions scored	148	149
Number of field testers	174	140
Mean number of correct responses	65.1	68.4
Standard deviation	14.3	17.5
Range of correct responses	30 - 103	30 - 115
Standard error of measurement	5.4	5.4
Hoyt's ANOVA - reliability	0.86	0.9

asthma/COPD ($p < .05$, *t*-test). A breakdown of mean performance on each component of both exams is provided in Table 2.

The individual item analysis was performed on the 297 questions from both exams. Seventy-nine questions were dropped from Exam A: 28 for low discrimination (point biserial correlation coefficient), 16 for low difficulty score (questions too difficult), 31 for a combination of low difficulty and discriminating ability, and 4 for high difficulty scores. Seventy-five questions were dropped from Exam B: 30 for low discrimination, 22 for low difficulty scores, 16 for a combination of low difficulty and discrimination ability, 4 for high difficulty scores, and 3 for high difficulty rating and low discrimination. Out of the 297 questions, a total of 143 (48%) questions met the criteria for difficulty and discrimination. These questions were used to develop the final test instrument consisting of 100 questions. To determine which of the 143 questions would be included in the final 100-point test instrument, questions were again reviewed for item difficulty and discrimination. Questions within each exam subtest (i.e., pharmacology, medicinal chemistry, etc.) closest to the optimal difficulty (60%) were included. For items of equal difficulty rating, the question with the highest discrimination score was chosen. The final test instrument contains 100 questions, all with a point biserial correlation coefficient ≥ 0.2. Table 3 provides a summary of the point biserial correlation coefficient for all questions included in the final test instrument. Six percent of the questions in the final instrument are in the 30% difficulty range, 16% are in the 40% range, and the majority (78%) fall into the 50-80% range. The breakdown of the final 100-point test instrument is: 10 medicinal chemistry questions, 9 pharmacology questions, 9 pharmacokinetics questions, and 8 questions each from 9 therapeutic categories. The therapeutic category of seizures was eliminated from the final instrument, since not enough questions of appropriate difficulty and discrimination were available.

DISCUSSION

A valid and reliable 100-point comprehensive examination was developed. The validity of this examination was established through

TABLE 3

Point Biserial	Number of Questions
0.2 - 0.29	39
0.3 - 0.39	36
0.4 - 0.49	19
0.5 - 0.59	6

an internal review panel that assessed questions for content validity. Reliability of this exam was calculated using Hoyt's ANOVA, which provides a measure of test reliability through a method of internal consistency. Internal consistency refers to the extent to which each test item measures whatever the test is measuring. For example, if each test item is considered a sample test from the total domain, then the internal consistency is equivalent to the average correlation between all pairs of items. For examinations designed to assess performance, the reliability coefficient should be 0.85 or higher. The Hoyt estimates of reliability for Exam A and Exam B were 0.86 and 0.9, respectively. In addition, the standard error of measurement–an index of quality similar to a reliability coefficient–was 5.4 for both exams.

The comprehensive examination was developed from an initial pool of 324 questions. From the 300 questions that were field tested, only 48% met the appropriate criteria for difficulty and discrimination. This is consistent with the expected dropout rate of 50-60% cited by other investigators developing test instruments (4). These results may be extrapolated to suggest that on any given multiple-choice exam within our college, approximately 50% of the examination questions may be of inappropriate difficulty or discrimination. This suggests that faculty members may benefit from formalized instruction in writing examination questions.

The overall performances on Exam A and Exam B were not significantly different. Initially, the mean scores of 65.1 (44%) and 68.4 (45.9%), respectively, seemed surprisingly low. However, it

seems reasonable to expect scores on field-tested exams to be somewhat lower, since participants are not familiar with faculty members' style of question writing. An additional factor that may help explain the scores is that students were not told to study for this examination. There were also no incentives provided for good performance or penalties given for poor performance. The mean scores on the field test were consistent with preliminary scores on a standard competency examination developed by Pray and Popovich for doctor of pharmacy students (4). The authors anticipate that the mean score on the final test instrument will improve, since questions of poor difficulty and discrimination have been excluded.

The final 100-point comprehensive examination will be administered to fifth-year pharmacy students at Wayne State University prior to the initiation of experiential training starting in the fall of 1991. Students and clinical preceptors will be provided with a breakdown of performance by subtests to identify individual areas of weakness. During experiential training, preceptors will require students to pick up case studies in their identified area(s) of weakness whenever possible. Required reading material may also be assigned at the discretion of the preceptor. This comprehensive examination will provide a mechanism by which students can strengthen their areas of weakness prior to graduation. Keeping this examination current will require at least yearly review of all items for content validity.

CONCLUSION

A valid, reliable comprehensive examination has been developed and will be administered to pharmacy students in their last professional year preceding experiential training. The results of this examination will identify student weaknesses for preceptors, allowing tailoring of the experiential program to individual needs. This will enhance the educational experiences of individual students and allow them to improve upon their identified weaknesses prior to graduation.

REFERENCES

1. Fassett WE, Campbell WH. Basic pharmaceutical sciences examination as a predictor of student performance during clinical training. Am J Pharm Educ 1984;48:239-42.

2. Gehres RW, Mergener MA, Sarnoff D. Comprehensive placement examination for sequencing pharmacy students into experiential courses and for determining appropriate remediation. Am J Pharm Educ 1989;53:16-9.

3. Kimberlin CL, Hadsall RE, Gourley DR, Benedict LK. Predicting success of pharmacy students in basic science and clinical clerkship courses. Drug Intell Clin Pharm 1983;17:297-301.

4. Pray WS, Popovich NG. The development of a standardized competency examination for doctor of pharmacy students. Am J Pharm Educ 1985;49:1-9.