

TEACHERS' TOPIC

A Five-Year Evaluation of Examination Structure in a Cardiovascular Pharmacotherapy Course

Anne Schullo-Feulner, PharmD, Claire Kolar, PharmD, Kristin K. Janke, PhD

University of Minnesota College of Pharmacy, Minneapolis, Minnesota

Submitted January 22, 2015; accepted June 30, 2015; published September 25, 2015.

Objective. To evaluate the composition and effectiveness as an assessment tool of a criterion-referenced examination comprised of clinical cases tied to practice decisions, to examine the effect of varying audience response system (ARS) questions on student examination preparation, and to articulate guidelines for structuring examinations to maximize evaluation of student learning.

Design. Multiple-choice items developed over 5 years were evaluated using Bloom's Taxonomy classification, point biserial correlation, item difficulty, and grade distribution. In addition, examination items were classified into categories based on similarity to items used in ARS preparation.

Assessment. As the number of items directly tied to clinical practice rose, Bloom's Taxonomy level and item difficulty also rose. In examination years where Bloom's levels were high but preparation was minimal, average grade distribution was lower compared with years in which student preparation was higher.

Conclusion. Criterion-referenced examinations can benefit from systematic evaluation of their composition and effectiveness as assessment tools. Calculated design and delivery of classroom preparation is an asset in improving examination performance on rigorous, practice-relevant examinations.

Keywords: examinations, ARS, assessment, multiple-choice questions, case-based learning

INTRODUCTION

Accreditation standards in higher education are requiring increased accountability for educational outcomes. Specifically, the Accreditation Council for Pharmacy Education (ACPE) mandates assessment activities to collect information about the attainment of desired student learning outcomes.¹ Historically, the mainstay of any curricular student learning assessment plan has been written, multiple-choice examinations. As a result, colleges and schools of pharmacy must be assured of examination effectiveness as a measure of assessing authentic learning outcomes. Scholarship provides guidance on examination design and delivery such as alignment of examination items with objectives, effects of item writing guidelines, and faculty committees to develop progress examinations.²⁻⁵ Other research includes the benefits of computer-based testing, the process of transitioning to electronic examinations, and the use of audience response systems (ARS) for testing.^{3,6-8} These design and delivery issues are relevant, but it is also necessary to evaluate

examination composition and to assess the effectiveness of an examination as an assessment tool.

Preparing students for examinations is also important, and ARS may be a helpful tool. Audience response systems are a formative assessment strategy providing instructors and students with immediate feedback and the opportunity for improved understanding via debriefing on item answers.⁹⁻¹¹ Although ARS appears to increase learning,^{12,13} little information is available on what types of ARS questions can help improve examination performance.

This article discusses how, in a cardiovascular pharmacotherapy course, an examination was created, how students were prepared for it, and how both have evolved over 5 years. In addition, the article explores the analyses performed on the examination each year and evaluates the impact of the findings over time. Finally, this article examines what was learned from this longitudinal evaluation and how these findings can be applied to examination design and revision, regardless of content area.

Corresponding author: Anne Schullo-Feulner, PharmD, Pharmaceutical Care & Health Systems, University of Minnesota College of Pharmacy, 7-103 Weaver-Densford Hall, 308 Harvard St. SE, Minneapolis, MN 55455. Tel: 952-993-5319. Fax: 952-993-6813. E-mail: amsf@umn.edu

DESIGN

The University of Minnesota College of Pharmacy offers its doctor of pharmacy (PharmD) program on 2 campuses (approximately 165 students) in Minneapolis

and in Duluth. Didactic instruction is delivered on one campus and broadcast via interactive television to the other. As one of 4 key required courses in the curriculum focusing on therapeutics, Pharmacotherapy II is a 5-credit course consisting of 5 hours of classroom instruction per week for 15 weeks. It is a second-year course designed to teach the pathophysiology and pharmacotherapy of cardiovascular, endocrine, and gastrointestinal disorders, which are covered as 3 separate sections. The cardiovascular section is evenly divided into 3 units: hypertension/dyslipidemia, arrhythmias/heart failure, and ischemic heart disease (IHD). Each unit is taught by one faculty member and concludes with a cumulative, written, 20-multiple-choice item examination. The assessment design and evaluation process delineated here describes the cumulative examination covering the IHD spectrum, including chronic stable angina and acute coronary syndromes.

The content and teaching methodology for didactic instruction and assessment stem from a focus on preparing students for the clinical challenges encountered as medication use experts during patient care. Pharmacotherapy knowledge is taught through problem solving and critical thinking about patient cases taken directly from an inpatient telemetry (cardiology) unit of a medium-sized (450 bed) community hospital with an attached, outpatient cardiovascular clinic. Skills and knowledge chosen for didactic instruction are those employed at least weekly by clinical pharmacists (often referred to as “decentralized” or “nondispensing” pharmacists). Skills critical for medication safety are also covered. Learning objectives focus on creating and adjusting a medication regimen, evaluating doses, and monitoring for adverse drug events.

Audience response systems are used during didactic instruction to provide students with a sequence of challenges to help build on their foundational knowledge and experiences through trial and error. For example, an early ARS question might be: “68-year-old Mr. Jones has chronic stable angina. He has no complaints of angina today. His blood pressure and (heart rate) = 148/83 mmHg (72 beats per minute). How should you manage his condition today?” Response options might include: “Suggest adding a long acting nitrate to prevent pain; suggest adding a beta blocker to bring heart rate to goal; suggest adding both; suggest continuing present management.” This type of short, straight-forward scenario ensures students are able to apply guidelines and prepares them to tackle more gray areas of patient care. Roughly 40% of instruction is dedicated to ARS sessions comprised of more complex patient case vignettes and progressive medical challenges (ie, patients with 3-5 disease states

and 5-10 laboratory results). Students progress through Bloom’s Taxonomy as they analyze rational options and synthesize multiple knowledge sources.

Assessment in the IHD unit follows the same pedagogical principles as instruction (ie, application of knowledge to authentic clinical challenges). Students are given small stakes quizzes prior to the unit’s cumulative, 20-item, multiple-choice, case-based examination. Because items are constructed to mimic challenges of working pharmacists, the examination is primarily “criterion-referenced” (ie, student performance is measured relative to a set standard and is independent of the performance of others).¹⁴ Students are not graded “on a curve,” and the overarching instructional goal is to enable all students to achieve the necessary criteria for passing. Students who score less than 60% on the examination fall below the minimum threshold for passing, resulting in concern for their ability to adequately address common and/or critical cardiovascular drug therapy problems. This approach to testing is a contrast to norm-referenced testing where student performance is measured relative to one another, but does not provide information regarding proficiency.¹⁴

From 2010 through 2014, the examination-structuring goal was to have all items taken directly from authentic common and/or critical practice scenarios. In any given examination year, approximately 40% of items were taken from pharmacist experiences in the outpatient setting (cardiovascular specialty clinic), and 60% of items were taken from pharmacist experiences in the inpatient setting (telemetry unit). The material itself was taught over nine, 50-minute periods and contained 14 distinct learning objectives (Table 1). These learning objectives were operationalized during class through ARS questions focused on decision making and assessed in the examination. Figure 1 illustrates how examination structuring was a strategic result of practice-informed learning objectives. Examination composition deliberately varied in terms of the number of items with strategic similarity to those used during instruction.

To evaluate the examination as it was being used year to year, indicators of the examination’s effectiveness and the effect of exposing students to preparation questions in class were identified. Five evaluation parameters were monitored and recorded over the 5 years: (1) degree of cognitive complexity using Bloom’s Taxonomy and association to practice activities; (2) similarity of examination items to cases reviewed in class [preparation similarity classification (PSC)]; (3) point biserial correlation coefficient (PBS); (4) item difficulty (ID); and (5) examination grade distribution (EGD). Adjustments were made to the examination each year, based on results.

Table 1. Preparation Similarity Classifications that Match Audience Response System (ARS) Questions to Examination Items

Preparation Similarity Classification	Classification Definition	Example Learning Objective	Example ARS Question**	Example Exam Item**
1. Equivalent Question	The patient parameters may vary slightly, but the learning objective tested and decision are the same.	Initiate, prioritize, and/or discontinue ACE-I or ARB therapy to optimize guideline supported morbidity/mortality outcomes in patients with Ischemic Heart Disease.	“56 y.o. Mr. Smith is a CSA patient who no longer reports chest pain. Currently on: ASA, statin therapy and BB. PMH: CKD (macroalb-urea) and MI. BP(HR): 145/87(65), CrCl: 60 mL/min” A. **Start an ACE-I B. Start ISMN C. Increase metoprolol D. Start diltiazem	“JP is a 60 y.o. male CSA patient who reports that his angina symptoms are controlled. Currently on: ASA, statin therapy and BB. PMH: CKD and MI. BP(HR): 148/89(70), CrCl: 45 mL/min” A. **Start an ACE-I B. Start ISMN C. Increase metoprolol D. Start diltiazem
2. Parallel Question	The patient parameters are different. The learning objective tested and decision are the same.	Categorize a patient with an Acute Coronary Syndrome as at greater risk of bleed or thrombosis and identify patients that would benefit from UFH v LMWH v either	“Choose an anticoagulant for a 62 y.o. female with NSTEMI. PMH: hx GI bleed. Wt: 43 kg. CrCl: 25 mL/min” A. **UFH B. LMWH C. UFH or LMWH D. warfarin	“Choose an anticoagulant for JH, a 78 y.o. male with unstable angina. PMH: HLD. Wt: 180 kg CrCl: 105 mL/min” A. **UFH B. LMWH C. UFH or LMWH D. warfarin
3. Same Concept	The patient parameters and decision are different. The learning objective tested is the same.	Choose a guideline supported, patient specific, therapy to medically manage uncontrolled chest pain in patients with Chronic Stable Angina.	“50 y.o. CSA patient, Ms. Smith, has been using her sublingual nitroglycerin frequently for activities of daily living due to her chest pain. Currently on BB. BP(HR): 165/96(55)” A. CPM if 1 NTG handles pain B. Increase her BB dose C. Add LA NTG D. **Add amlodipine	“BN is a 77 y.o. woman with CSA complaining of dizziness when she stands up quickly and a curbed exercise routine due to daily chest pain. Currently on BB. BP(HR): 115/75 (58)” A. CPM if 1 NTG handles pain B. Increase her BB dose C. **Add LA NTG D. Add amlodipine

(Continued)

Table 1. (Continued)

Preparation Similarity Classification	Classification Definition	Example Learning Objective	ARS Question**	Example Exam Item**
4. No ARS Match	The patient parameters, learning objective tested, and decision are different.	Identify common or serious side effects, precautions, and monitoring strategies for all medications traditionally used in patients with Ischemic Heart Disease. Drug class examples include: BB, CCB, nitrates, ACE-I/ARBs, morphine, anti-platelets, anticoagulants, etc.	Not Applicable	<p>“Which medication monitoring statement is most correct?”</p> <p>A. If amlodipine is added, the patient’s potassium level should be monitored at each visit.</p> <p>B. If ISMN is added, care needs to be taken to ensure that nitrate blood levels remain consistent over each 24 hour period.</p> <p>C. If lisinopril is added, we would anticipate that the patient’s systolic blood pressure would remain the same at it was at baseline (prior to the addition).</p> <p>**D. If metoprolol is added, we should monitor the patient’s AV node conduction either via heart rate or PR interval.</p>

**Indicates the correct option/response

Medical Abbreviations: ACE-I=angiotensin-converting enzyme inhibitor, ARB=angiotensin receptor blocker, CSA=chronic stable angina, ASA=aspirin, BB=beta-blocker, PMH=past medical history, CKD=chronic kidney disease, MI=myocardial infarction, BP=blood pressure, HR=heart rate, CrCl=creatinine clearance, y.o.=year/s old, ISMN=isosorbide mononitrate, UFH=unfractionated heparin, LMWH=low molecular weight heparin, NSTEMI=non ST-elevation myocardial infarction, hx=history of, GI=gastrointestinal, wt=weight, kg=kilogram, mL=milliliter, HLD=hyperlipidemia, CPM=continue present management, NTG=nitroglycerin, LA=long acting, CCB=calcium channel blocker, AV=atrioventricular

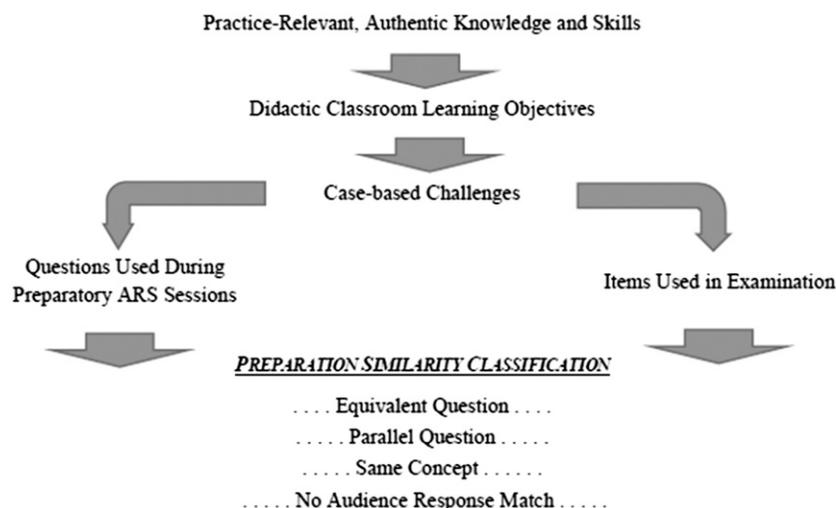


Figure 1. Examination Structuring Design as a Strategic Result of Practice-based Challenges and Didactic Classroom Learning Objectives.

For the analysis, each examination item was assigned one of the following 3 condensed levels of Bloom’s Taxonomy: recall (ie, knowledge and comprehension), application, or analysis (ie, analysis, synthesis, and evaluation).¹⁵ Two instructors independently assigned these ratings to each item. Ratings were then compared and discussed. Because the majority of pharmacist responsibilities in the care setting involve a minimum of applying knowledge and skills to an individual patient, for this analysis, Bloom’s Taxonomy level functioned as a measure of practice readiness.

Examination composition was deliberately varied each year in terms of the number of items with similarity to those used during ARS sessions. As the complexity, and thus Bloom’s Taxonomy level, of cases and items rose to meet the level of clinical practice, the instructor altered student preparation during class time to ensure instruction level matched assessment level. Specifically, the instructor classified and recorded the type of in-class preparation for each item as equivalent, parallel, same concept, or no match. Definitions and examples of these preparation similarity classifications are provided in Table 1.

Point biserial and ID were available from reports routinely provided by the university’s Office of Measurement Services and were monitored by the instructor over time. The point biserial indicates how well an item discriminates between higher-performing and lower-performing students. In general, the higher a PBS value, the more discriminating the item. However, most items should aim to have a PBS of ≥ 0.2 .^{16,17} The desired number of “strong” PBS items in an examination may vary depending on the purpose of the examination. In a criterion-referenced examination, it is possible for all students to get a perfect

score.¹⁸ For items intended to assess mastery of learning objectives and clinical practice competence, the PBS may be appropriately below a 0.2 level. Point biserial data is best interpreted in tandem with ID, which represents the percentage of students getting an individual item correct. With a criterion-referenced examination, the instructor hopes and anticipates the majority of students will do well rather than targeting a bell-shaped grade distribution.

To evaluate an item’s effectiveness in this cardiovascular criterion-referenced examination, the instructor set a goal item difficulty benchmark of 85% or above, while items with an ID of 60% or below were viewed with concern. These two ID benchmarks were used in combination with the PBS to evaluate items in 3 distinct ways: (1) When the ID was at or above 85%, the PBS might appropriately fall below 0.2, indicating that the majority of students successfully achieved that learning objective; (2) When the ID was 85% to 60% and the PBS was well above 0.2 (eg, 0.4), this indicated that an individual item was difficult, but in general, students who performed well on the examination also performed well on the item. This pair of conditions showed that while the item was difficult for students, it was also well aligned with its learning objective and classroom instruction; (3) When the ID of an individual item was less than 60%, the PBS was examined. If the PBS was below 0.2, the instructor considered whether the item itself was flawed and therefore not an accurate representation of student knowledge, or if the learning objective being tested was not adequately taught/prepared for in class. If the PBS was above 0.2, the item was considered difficult, but the top performers did well. Being a criterion-referenced examination, such items may have indicated the learning

objective being tested was not adequately taught or prepared for in class.

The final tool used to evaluate examination effectiveness each year was overall examination grade distribution, which was available from the instructor's gradebook. As a criterion-referenced examination, the overarching goal was for all students to achieve a passing score or better. However, a priority was placed on ensuring a high level of practitioner competency through rigorous assessment. As a result, it was recognized that passing scores weren't possible for all students. Student satisfaction data was taken from the anonymous, optional, end-of-semester course evaluation survey. This project was reviewed by the University of Minnesota Institutional Review Board (IRB) and found to be exempt.

EVALUATION AND ASSESSMENT

Five cumulative unit examinations were given over 5 years of instruction from 2010 to 2014. Each year, approximately 165 second-year pharmacy students, 108 at the Minneapolis campus and 57 at the Duluth campus, participated in ARS sessions and took the IHD unit examination. Each cumulative unit examination consisted of 20 case-based items in 2010 through 2013. The 2014 examination contained 17 items as a result of adding cardiovascular informatics material to the unit. In total, 97 examination items were used and evaluated.

Bloom's application and analysis levels were dominant in all 5 examinations. As planned by the instructor, there was a shift to more analysis-based items over time (Figure 2). In addition, from 2010 to 2011, the examination's composition changed from 70% of items being

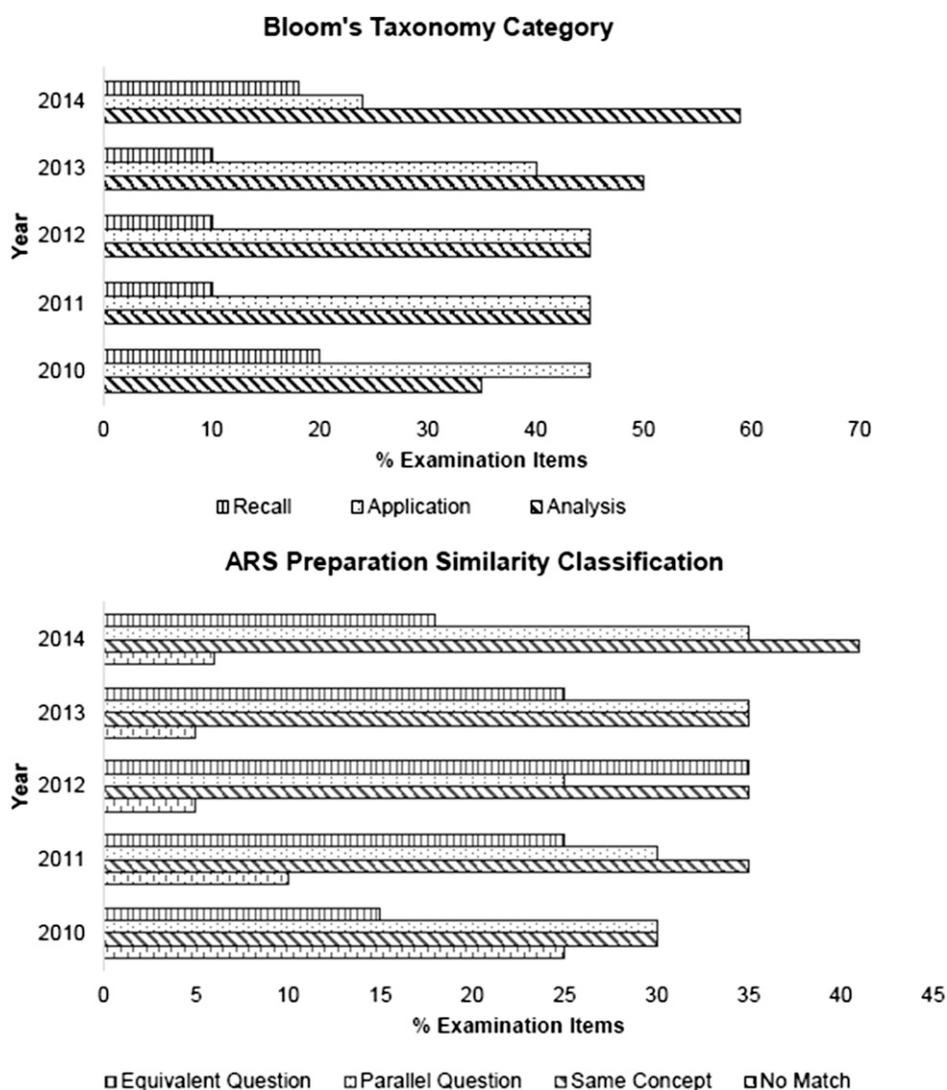


Figure 2. Distribution of Items by Bloom's Taxonomy Category and ARS Preparation Similarity Classification.

directly aligned with practice activities to 90% being aligned with practice activities. From 2013 forward, 100% of items were directly aligned with frequent practice activities and/or critical knowledge and skills.

Moving from the 2010 examination to the 2013 and 2014 examinations, the cognitive function needed to correctly arrive at clinical solutions correlated increasingly to higher Bloom's Taxonomy levels. Students needed more concrete practice making such decisions, so the instructor shifted from using items with no ARS preparation to using items practiced by students in class (Figure 2).

Point biserial and ID could not be compared across the preparation similarity classifications (PSC) or Bloom's Taxonomy levels. The work with this examination was not structured as an experiment with controlled numbers of items allocated to Bloom's Taxonomy or to

preparation classifications. In addition, various questions were repeated at different points, complicating analysis. However, observations were made from the descriptive data. The distribution of examination items by PBS and ID is available in Figure 3. While 2010 had the highest number of items with a PBS below 0.2, the ID data for that year indicated that more than 81% of students answered a large percentage of items correctly. When seen in context with the ID data for that year, the PBS data is appropriate.

Many items on the 2010 examination were not difficult, so it is not surprising that discrimination between high and low performers on certain items was lower. In 2011, a shift was seen toward higher point biserials coupled with lower ID (ie, fewer students did well). This undesirable shift was attributed to a higher number of

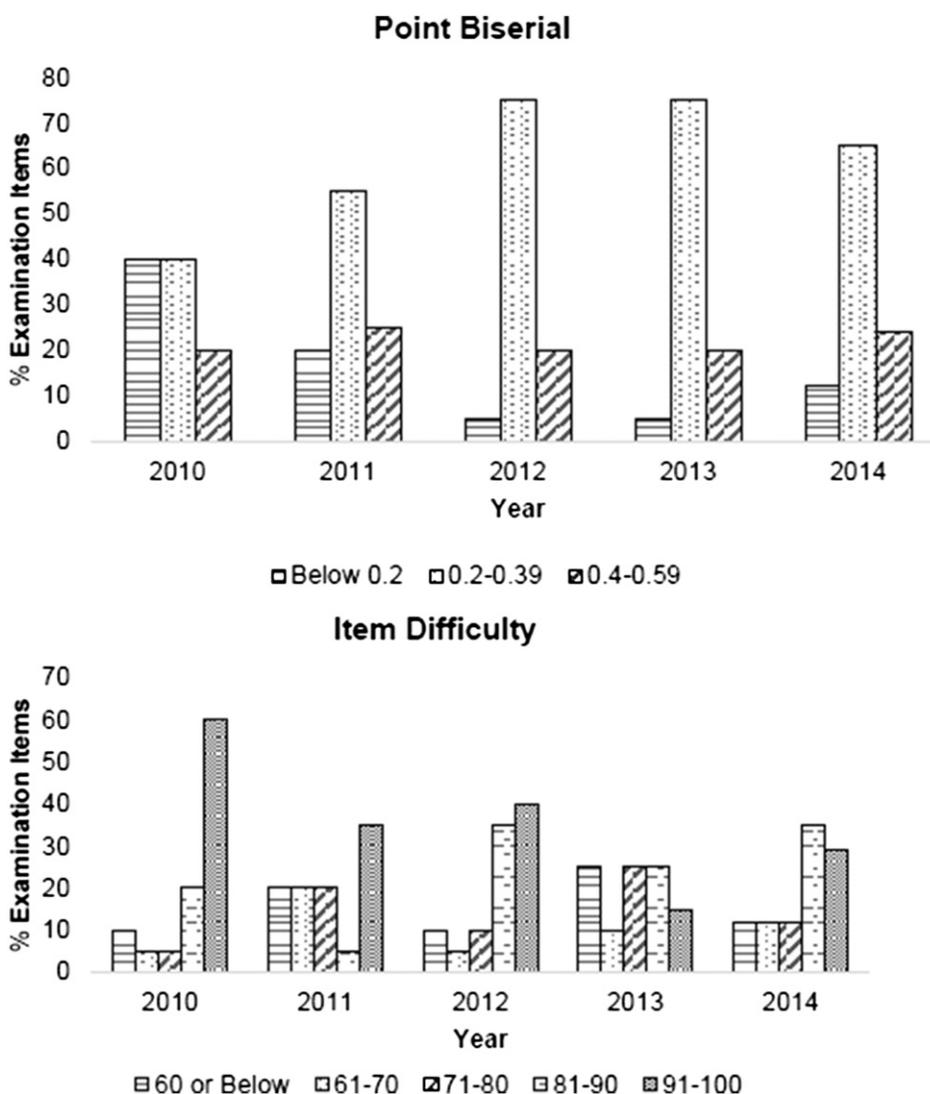


Figure 3. Point Biserial and Item Difficulty.

2011 items being directly tied to practice readiness and direct patient care scenarios. Accordingly, Bloom’s Taxonomy data from 2010 to 2011 demonstrated an increase in items at the levels of application and analysis. Thus, the instructor modified classroom-based ARS cases to better prepare students. This strategy was successful, as years 2012, 2013, and 2014 all show more balanced ID data coupled with an increased number of items at a Bloom’s Taxonomy level of analysis. While items where less than 70% of students answered accurately were lower than in previous years, the PBS remained high overall.

The examination grade distribution indicated that the majority of students did well. As shown in Figure 4, as items shifted to higher levels of Bloom’s Taxonomy, and student preparation varied, the grade distribution for the examination shifted. In 2010, when items demonstrated lower levels of Bloom’s Taxonomy and less connection to practice, students did well, with over 80% of the class scoring 80% or more on the examination. When items were better tied to practice readiness (and higher Bloom levels) in 2011, only 50% of students scored higher than 80%. However, once ARS-driven, in-class preparation strategies were put in place in 2012, 2013, and 2014, between 5 and 12 students (out of 165) earned a 100% score on the high Bloom-level examination. On the 2014 examination (an assessment on which 100% of the questions were practice relevant and more than 80% of items were written at a Bloom’s Taxonomy level of application or higher), 65% of students scored at or above 80%. This level of practice relevance, coupled with advanced cognitive requirements and a favorable grade distribution could be interpreted as a sign of effective teaching.

Table 2 delineates the differences between the ARS preparation similarity classifications, in terms of ID and

PBS. The average ID was the highest for “equivalent questions,” indicating these were the least difficult for students to answer. Yet, the average PBS for these items was not below 0.2, and not all students answered correctly. The “equivalent question” preparation technique is used for skills used frequently or that are critical safety issues in practice. The most difficult items for the students were in the “same concept” classification, which also had the highest average PBS. The PBS for each category was above 0.2, indicating items in all 4 categories distinguished between high and low performers on the examination.

Once the initial investment in ARS software and hardware was made, it required little else in terms of resources aside from instructor effort to write questions. In course evaluations, students showed appreciation for the use of both patient cases, as well as for ARS. On the 2014 postsemester course evaluation, 48% of 163 students responded to 2 questions. The use of patient cases in the cardiovascular section received an average rating of 6.1 (SD=0.9), indicating a student response of “very useful” on the 7-point Likert scale (1=not at all useful to 7=extremely useful). Similarly, the use of ARS received an average rating of 5.6 (1.2), indicating a student response of “effective” on the 7-point Likert scale (1=very ineffective to 7=very effective).

DISCUSSION

Over the 5-year time period, this cardiovascular examination composition shifted to include more items at higher levels of Bloom’s Taxonomy, and 100% of items directly assessed practice-relevant activities. Although multiple-choice examinations are traditionally written at a recall level,¹⁹ deliberately creating a larger proportion

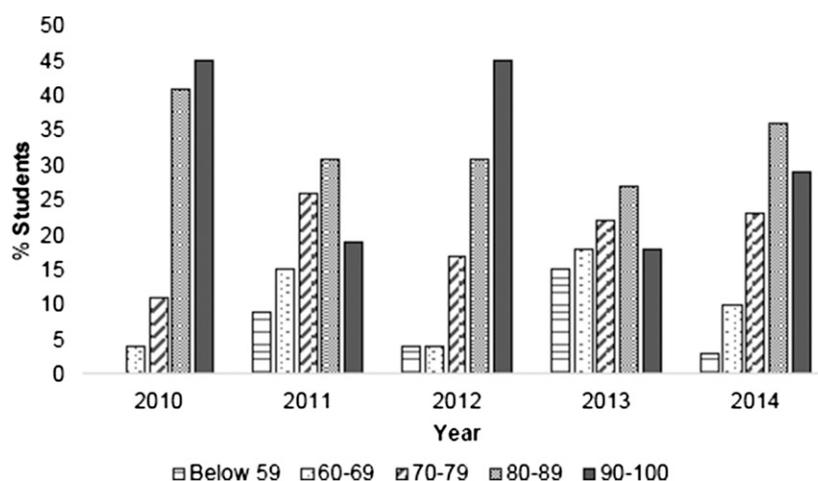


Figure 4. Examination Grade Distribution Over 5 Years.

Table 2. Average 2014 Item Difficulty and Point Biserial by Similarity Classifications

Item Prep Category	Equivalent Question (n=3)	Parallel Question (n=6)	Same Concept (n=7)	No Match (n=1)
Average Item Difficulty	91.67%	85.17%	70.43%	84%
Average Point Biserial	0.23	0.29	0.34	0.31

of application and analysis questions can better prepare students for the critical thinking and problem solving needed in practice settings. However, while this analysis reports higher Bloom’s levels and better practice relevance over time, fundamentally, Bloom’s Taxonomy only provides information on the cognitive processing required, not necessarily an item’s applicability to daily practice (eg, frequency of use, criticality of information). With the increased interest in advanced pharmacy practice experience (APPE) readiness and practice readiness, it may be helpful for instructors to consider a systematic means for describing this connection. Examination support technology often makes it possible to “tag” items with terms (eg, practice readiness categories) to aid in analysis.

Often, students do not initially perform well on practice-authentic ARS cases. This can create frustration during class, but it engages students and gives them immediate feedback on their ability to apply knowledge to patient cases. To allow students to become successful at a high degree of practice readiness, varying proportions of equivalent, parallel, and same concept questions were used in ARS sessions as preparation for the cardiovascular examination. However, examination preparation is not the only use for item classification. Williams et al categorized final examination items based on their relationship to midterm examination items— identical question, same emphasis, moderately different emphasis, and different emphasis.²⁰ After completing an activity analyzing items missed on the midterm examination, they used the categories to determine which type of item yielded the best student performance. They found students performed best on final examination items with the same emphasis as midterm examination items. After 5 years of ARS preparation analysis and adjustment in this course, most items were either parallel to, or used the same concept as, those used during in class preparation. Such items tested students’ ability to apply their knowledge to different situations, an essential skill in the practice setting. Less frequently utilized were items equivalent to preparation questions, although such an educational strategy would make sense for practice skills used frequently or that are critical for patient safety. Even when this strategy was employed, not all students answered correctly. In general, this may speak to the difficulty of items at Bloom’s levels of

application or higher; it also may identify students who have a significant deficit and need remediation. Our data set suggests using a preparation similarity classification system for examination items and preparation questions may enable instructors to test high level Bloom’s cognitive functions and rigorous practice-relevant skills while maintaining ID and PBS levels.

While PBS and ID are often readily available in examination administration reports, they frequently are used for their immediate relevance in identifying poor items that must be dropped on the current examination. However, as demonstrated in this course evaluation, these data also can be used to aid in decision-making for future examinations. Specifically, discussion of the desired proportions of higher and lower PBS items is useful. If all items are highly discriminating, there may be an unintended effect on total scores by eliminating questions whose purpose is to assess critical competencies for passing (ie, ensuring the effective and safe practitioner level). A proportion of lower PBS and difficulty items may be appropriate for items testing very frequent or critical practice skills where the majority of students do well and demonstrate mastery of objectives.

While the longitudinal monitoring and continuous quality improvement conducted in this study benefited the instructor, students, and course, it also revealed considerations that may be useful to other course instructors and those administering curricular level examinations. Based on this longitudinal analysis, the following 7 examination-structuring guidelines are proposed: first, as a fundamental starting place, an agreement on the approach to testing (ie, criterion-referenced vs. norm-referenced) is needed. Assessment leaders in pharmacy recommended that criterion-referenced testing is best practice and should be a goal.¹⁹ In addition, criterion-referenced examinations are appropriate in medicine.¹⁸ When evaluating criterion-referenced examinations, the determining factor for item inclusion does not need to be based on difficulty or discriminating ability, but rather on how well the item accurately reflects the desired competency.²⁰ Second, and equally important, items should be tied directly to desired skills and knowledge (in this case pharmacy practice). Skills that are frequently used and/or critical should be the primary focus. Methods for categorizing items, including practice relevance, are encouraged. Third, Bloom’s Taxonomy levels can be used

to ensure assessment of the desired level(s) of cognition. Assessments of health care competencies should aim for a level of application or higher. Fourth, classroom preparation should take into account the Bloom's level required and skill acquisition difficulty. Preparation should be deliberate and vary accordingly. In pharmacy education, the use of parallel or same concept preparation questions can aid students in applying knowledge to different situations – a critical practice skill. Fifth, when teaching critical thinking and decision making, consideration should be given to the use of ARS as a tool for examination preparation. Sixth, PBS and ID should be assessed in tandem and tracked over time. For items with a high degree of difficulty, classroom preparation should be examined. In criterion-referenced assessment, low point biserials may be appropriate and desired for items assessing mastery of basic competencies. Finally, the overall grade distribution should be monitored over time with the goal of moving the class as a whole toward 100% accuracy on all mastery items.

This work has a number of limitations, including the lack of an experimental design and the small number of items. Future research should carefully control item numbers and categories of items over time, allowing for more sophisticated analysis.

SUMMARY

This project involved the evaluation of a 20-item multiple choice, practice-aligned examination using Bloom's Taxonomy and preparation similarity classification. Examination effectiveness as an assessment tool of student learning was evaluated using point biserial correlation coefficients, item difficulty and examination scores. These data were monitored over 5 years, and changes in instruction were made to optimize student learning. From 2010 to 2014, examination items gradually became 100% tied to authentic clinical practice-based skills, with a correlated rise in cognitive complexity. The use of high-level, challenging ARS cases and questions in class was used to encourage practice readiness, to teach complicated material and decision making, and to improve the ability of examinations to evaluate student learning.

REFERENCES

1. Accreditation Council for Pharmacy Education. Accreditation standards and guidelines for the professional program in pharmacy leading

- to the doctor of pharmacy degree. 2007. <https://www.acpe-accredit.org/pdf/FinalS2007Guidelines2.0.pdf>. Accessed September 3, 2014.
2. Wittstrom K, Cone C, Salazar K, Bond R, Dominguez K. Alignment of pharmacotherapy course assessments with course objectives. *Am J Pharm Educ.* 2010;74(5):Article 76.
 3. O'Brocta R. Computer testing to document student achievement of learning outcomes. *Am J Pharm Educ.* 2013;77(10):Article 226.
 4. Pate A, Caldwell DJ. Effects of multiple-choice item-writing guideline utilization on item and student performance. *Curr Pharm Teach Learn.* 2014;6(1):130-134.
 5. Medina MS, Britton ML, Letassy N a., Dennis V, Draugalis JR. Incremental development of an integrated assessment method for the professional curriculum. *Am J Pharm Educ.* 2013;77(6):Article 122.
 6. Bussieres J-F, Metras M-E, Leclerc G. Use of Moodle, ExamSoft, and Twitter in a first-year pharmacy course. *Am J Pharm Educ.* 2012;76(5):Article 94.
 7. Pawasauskas J, Matson KL, Youssef R. Transitioning to computer-based testing. *Curr Pharm Teach Learn.* 2014;6(2):289-297.
 8. Kelley K a., Beatty SJ, Legg JE, McAuley JW. A progress assessment to evaluate pharmacy students' knowledge prior to beginning advanced pharmacy practice experiences. *Am J Pharm Educ.* 2008;72(4):88.
 9. Divall M V, Alston GL, Bird E, et al. Special article: A faculty toolkit for formative assessment in pharmacy education. *Am J Pharm Educ.* 2014;78(9):Article 160.
 10. Cain J, Robinson E. A primer on audience response system : Current applications and future considerations. *Am J Pharm Educ.* 2008;72(4):Article 77.
 11. Cain J, Black EP, Rohr J. An audience response system strategy to improve student motivation, attention, and feedback. *Am J Pharm Educ.* 2009;73(2):Article 21.
 12. Liu FC, Gettig JP, Fjortoft N. Impact of a student response system on short- and long-term learning in a drug literature evaluation course. *Am J Pharm Educ.* 2010;74(1):Article 6.
 13. Slain D, Abate M, Hodges BM, Stamatakis MK, Wolak S. An interactive response system to promote active learning in the doctor of pharmacy curriculum. *Am J Pharm Educ.* 2004;68(5):Article 117.
 14. Glaser R. Instructional technology and the measurement of learning outcomes: *Am Psychol.* 1963;18(8):519-521.
 15. Tiemeier AM, Pharm D, Stacy ZA, Burke JM. Using multiple choice questions written at various Bloom's Taxonomy levels to evaluate student performance across a therapeutics sequence. *Inov Pharm.* 2011;2(2):Article 41.
 16. Understanding the item analysis report. *Univ Minnesota Off Meas Serv.* http://oms.umn.edu/fce/understanding_results/itemanalysis.php#reliability. Accessed December 19, 2014.
 17. Tavakol M, Dennick R. Post-examination analysis of objective tests. *Med Teach.* 2011;33(6):447-458.
 18. Ricketts C. A plea for the proper use of criterion-referenced tests in medical assessment. *Med Educ.* 2009;43(12):1141-1146.
 19. Williams AE, Aguilar-Roca NM, Tsai M, Wong M, Beaupré MM, O'Dowd DK. Assessment of learning gains associated with independent exam analysis in introductory biology. *CBE Life Sci Educ.* 2011;10(4):346-356.
 20. Popham WJ, Husek TR. Implications of criterion-referenced measurement. *J Educ Meas.* 1969;6(1):1-9.