Turning a Problem into a Solution: Instructional Benefits of Using ExamSoft to Map Multiple-Choice Responses

Dan Cernusca
North Dakota State University

David M. Scott
North Dakota State University

Daniel L. Friesner
North Dakota State University

In this paper, we propose a simple method to minimize the effects of each of the previously mentioned issues associated with the design and administration of assessments that use multiple-choice items. We analyzed multiple-choice questions with possible responses to express fundamentally different thought processes and/or knowledge. A case study built around an introductory health care course shows how a specialized assessment software platform helped map assessment questions back to curricular outcomes identifying where the curriculum was effective, and where needed improvement. Our results indicate that incorrect responses were more informative in identifying and addressing curricular gaps than the correct responses.

BACKGROUND AND MOTIVATION OF THE STUDY

Most health professions programs are required by external accrediting bodies to provide evidence demonstrating that each skill, ability or component of knowledge delivered in the program adequately prepares students to successfully pass licensure and/or certification exams and to ultimately practice competently within that profession. As a result, many accredited health professions programs utilize curricular maps to tie course learning objectives, and the assessments that measure those objectives, back to the program's curricular outcomes. While many of these course-specific objectives require higher order skills (synthesis, analysis, etc.), other objectives require a baseline knowledge of facts and comprehension also known as "remembering" and "understanding"(Anderson et al., 2001).

The Accreditation Council for Pharmacy Education (ACPE) clearly defines the standards all professional Doctor of Pharmacy (PharmD) programs need to meet to maintain their accreditation (ACPE, 2015). The current 25 ACPE standards along with their associated key and assessment elements and required documentation ensure both practice and professional readiness of PharmD graduates.

A clear focus of ACPE standards is the importance of valid and reliable assessment mechanisms that are able to provide instructors and educational organizations with a clear picture of both the strength and

weaknesses of the instructional process. Assessment outcomes are the critical elements in identifying the adequacy of current educational resources to be in full compliance with ACPE standards.

Instructors teaching foundational courses, especially those faculty members who teach large lecture PharmD courses, often use multiple-choice exams as an efficient means to assess student learning. Because multiple-choice questions typically have only one correct answer, it is also relatively straightforward to use this type of assessment items in curricular assessment activities. Students who correctly answer specific exam questions provide evidence (as traced back through the curriculum map) demonstrating mastery, or movement towards mastery of specific programmatic goals. Students who answer the question incorrectly display a lack of competency, although the source of the competency gap is often unclear.

The difficulty in using multiple-choice assessments lies in the interpretation of a correct or incorrect response. The educational assessment literature has identified a number of potential problems with standardized multiple-choice tests, including, but not limited to, teaching to the test, poorly written questions, test item bias, and placing too few questions per knowledge construct on the exam (e.g. Burton, 2002; Downing, 2002a). Most of these issues deal with errors on the part of faculty that struggle to integrate the assessment in the design of the entire course. That is, rather than making assessment development a part of the early course design (Morrison, Ross, Kalman, & Kemp, 2011) faculty often scramble to put together assessment items at the end of a given section or course, often losing the connection between what was actually taught and what is assessed with these items.

Other test flaws are primarily the result of the multiple-choice design. Two of the most common problems with multiple-choice tests (or any tests with a discrete number of pre-defined responses) are "guessing the right answer" and using strategies to deduce the correct answer without actually knowing the answer; called "testwiseness"(Bereby-Meyer, Meyer , & Flascher, 2002; Downing, 2002a, 2002b, 2003). Assessments that fail to account for one or more of these flaws, will report upwardly biased learning outcome metrics and enable marginally competent students to complete the course with apparent success.

Objective of the Study

In this paper, we propose a simple method to minimize the effects of each of the previously mentioned issues associated with the design and administration of assessments that use multiple-choice items. However, the instructor intentionality predicates the effectiveness of this method. That is, when an instructor designs a specific multiple-choice question, each possible response (both correct and incorrect), must capture fundamentally different thought processes and/or knowledge. This way, each time a student provides a specific response, whether correct or incorrect, the student divulges unique information about her or his state of knowledge and his or her learning process. The correct response certainly indicates where gains in learning occurred. Equally important is that incorrect responses identify learning gaps, and provide clues about why learning did not occur.

Moreover, with proper, premeditated question design and the use of and exam administration software that allows to randomize questions and responses, it is possible reduce the likelihood of guessing and testwiseness. This provides the instructor with greater certainty that correct responses truly captured learning gains while incorrect responses identified knowledge gaps. A software such as ExamSoft® (http://learn.examsoft.com/) also allows the instructor to map each assessment question to the appropriate curricular outcomes. In doing so, it is possible to identify both sides of assessment: where the curriculum is effective, and where it requires improvement. The ability to move the analysis to curricular level adds a new level of assessment complexity, when compared to simply identifying a gap in knowledge acquisition.

This study's major objective is to provide a case study illustrating how intentional multiple-choice question design, coupled with the exam administration software ExamSoft®, can be used to assess where and why learning does occur or not. In addition, the proposed process provides a methodological basis for much richer assessment of learning, and ultimately an assessment of learning outcomes at the curriculum level.

Setting and Design

As an illustrative example, this case study focuses on the implementation of intentional question design and exam administration software ExamSoft® in an "Introduction to U.S. Health Care Systems" course taught in the Doctor of Pharmacy (PharmD) program at North Dakota State University. The program is accredited by the Accreditation Council for Pharmacy Education (ACPE) that requires detailed curricular mapping and assessment of student learning as key accreditation criteria (ACPE, 2015). The Introduction to U.S. Health Care Systems course is required of all students and its' offering for the first year (P1) PharmD students. This case study will cover these two semesters for the 2016 academic year. As such, it is subject to ACPE curricular mapping and assurance of learning requirements. The enrollment in the course typically ranges from 81 to 85 students per year.

The multiple-choice questions identified for this case study are part of the third exam of the Introduction to U.S. Health Care Systems course. The textbook for the course provides a general perspective of the U.S. health care system that would be appropriate for a variety of health professions majors. This section of the text covers general information about the organization of the U.S. system, with more specialized content (for example, health care financing, health economics, health services research, health disparities and culture, global health, etc.) covered during the remainder of the semester. Consistent with these themes, we identified for this analysis two questions that are applicable to a variety of health professions. The first exam question used in this case study is the following (with the correct answer highlighted in bold face font):

Which of the following is (are) true?

A. Most beneficiaries do not pay a premium for Medicare Part A.

B. Most beneficiaries do not pay a premium for Medicare Part B.

C. Payroll taxes are a major source of funding for Medicare Part A.

✓D. A and C

E. A, B and C

For this question, answer "D" is the correct one, as it indicates that students recalled what groups the instructor covered in various sections of Medicare. For those that chose either "A" or "C" it shows that they know and understood one part of the benefit, but not both of them. For those that chose "B," students did not comprehend either the information provided in the textbook, another assigned reading "Medicare and You for 2016", did not attend class, or did not grasp the information presented in class. Since this information is rather basic to the understanding of Medicare, the instructor expected that most of the students should do well on this item.

The second exam question used in this case study is the following:

Global trends from 1990 to 2010 show the following causes of mortality

A. Non-communicable causes rose and communicable, neonatal and nutrition causes rose

✓B. Non-communicable causes rose and communicable, neonatal and nutrition causes decreased

C. Non-communicable causes decreased and communicable, neonatal and nutrition causes rose

D. Non-communicable causes decreased and communicable, neonatal and nutrition causes decreased

This item is a global health question that assesses the mortality (death rates) over a 20-year period (1990-2010). It examines the influence of several trends such as the prevalence of communicable (infectious) diseases and the prevalence of chronic diseases (non-communicable). The decreased prevalence of neonatal and nutrition causes of conditions and diseases demonstrates the impact that public health has made worldwide. It examines the impact of public health measures (prevention, monitoring, control measures, nutrition, maternal and child health care) have made. While clinical medicine (diagnosis, treatment, management, antibiotics) has influenced the prevalence of disease, the impact is not at the same level that public health has made. For instance, public health's impact on infant mortality rates on a global basis has been notable in these areas. While the overall trend is downward, there are wide swings from country to country. Generally the developed countries (e.g., USA, United Kingdom, Canada, Western European countries), and developing countries (e.g., Mexico, Brazil, Russia, Eastern

European countries) have somewhat better rates than those in underdeveloped countries (e.g., Afghanistan, Congo, Kenya). Therefore, selection of choice "B" indicates that the student grasped the right concepts and trends. If the student chose "A", it showed they did not know that only chronic diseases rose and the three others have decreased globally. If the student chose "C", they failed to know that non-communicable diseases have increased. Finally, if the student chose "D" they failed to know that the basic changes have occurred, and that non-communication diseases have increased while communicable diseases have decreased.

It is important to emphasize that closing the students' knowledge gap requires more than addressing the incorrect response on the given exam item. Rather, each incorrect response provides an opportunity for the instructor to delve into the root etiology underlying that gap. Put differently, the instructor must refocus on the underlying rationale for a specific element of the U.S. health care system's regulatory structure. With regard to the first question in this case study, it is insufficient to remind students that option "*B*" is incorrect, and options *"A"* and *"C"* are both correct. Rather, it is vital to explain *how* and *why* premium payments relate to coverage, which in turn relate to (and are driven by) underlying epidemiologic and socio-economic considerations among Medicare enrollees. Only then will students truly grasp the rationale for requiring the payment of Medicare Part B premiums.

Assessment Analysis Tools and Results

All professional pharmacy students at North Dakota State University are required to purchase an Apple iPad with specific memory specifications. This facilitates electronic note taking, remote access to recorded lectures, and electronic exam administration. With regard to this last point, the exams that are the focus of this analysis was administered to students electronically using SoftTestM®, the iPad examination application in the ExamSoft® assessment platform. The ExamSoft® assessment platform allows both question and response option randomization, and ties each question to specific curricular outcomes (i.e., course objectives, ability-based outcomes, taxonomy levels). In addition, ExamSoft® provides a variety of reports on single or longitudinal assessment results. Because the primary objective of this case study is to illustrate a simple means to extract useful information mainly from incorrect responses (that would be accessible to any faculty, regardless of statistical, instructional design and psychometric background), the ExamSoft® reports will be used as the primary means of information extraction and analysis.

These data analysis tools in ExamSoft® allow the instructor not only to evaluate the quality of specific assessment items but also to reflect on the quality of the instructional process that generated the assessed results. In addition, ExamSoft® has several security features starting from full device "lock down" for closed-book exams to access to the device but no access to the internet resources for open-book exams. For proctored exams, these features significantly or totally reduce the propensity for students to use the technology inappropriately.

Evaluating Individual Assessment Items

One important tool to review the quality of multiple-choices items offered by ExamSoft® is the Item Analysis report. This report provides a wealth of information, including (but not limited to) student response distributions, item difficulty indices and item discrimination indices for both the question itself and its individual choices. Figure 1 illustrates an ExamSoft report provided for the two test questions analyzed in this case study for the fall 2016 administration of the assessment.

FIGURE 1
ITEM ANALYSIS OUTPUT FOR MULTIPLE-CHOICE EXAM ITEMS ADMINISTERED IN AN "INTRODUCTION TO U.S. HEALTH CARE SYSTEMS" COURSE.

Question #	Correct Responses			Disc. Index	Point Biserial	Correct Answer	Response Frequ				
	Diff(p)	Upper	Lower				A	B	C	D	E
Description: Financing											
43	0.80	95.65%	65.22%	0.30	0.26	D	1	0	5	*68	11
Questi... Rev: 241 / 3				-	-	% Selected	1.18	0.00	5.88	80.00	12.94
-	-	-	-	-		Point Biserial (rpb)	-0.17	0.00	-0.15	0.26	-0.15
-	-	-	-	-		Disc. Index	-0.04	0.00	-0.09	0.30	-0.17
-	-	-	-	-	-	Upper 27%	0.00	0.00	0.04	0.96	0.00
-	-	-	-	-	-	Lower 27%	0.04	0.00	0.13	0.65	0.17

Q: Which of the following is (are) true?
A: Most beneficiaries do not pay a premium for Medicare Part A.
B: Most beneficiaries do not pay a premium for Medicare Part B.
C: Payroll taxes are a major source of funding for Medicare Part A.
* **D:** a and c
E: a, b and c

Description: Developments											
30	0.46	69.57%	30.43%	0.39	0.29	B	5	*39	24	17	-
Questi... Rev: 259 / 3				-	-	% Selected	5.88	45.88	28.24	20.00	-
-	-	-	-	-		Point Biserial (rpb)	-0.14	0.29	-0.26	0.01	-
-	-	-	-	-		Disc. Index	-0.13	0.39	-0.26	0.00	-
-	-	-	-	-	-	Upper 27%	0.00	0.70	0.09	0.22	-
-	-	-	-	-	-	Lower 27%	0.13	0.30	0.35	0.22	-

Q: Global trends from 1990 to 2010 show the following causes of mortality
A: Non-communicable causes rose and Communicable, neonatal and nutrition causes rose
* **B:** Non-communicable causes rose and Communicable, neonatal and nutrition causes decreased
C: Non-communicable causes decreased and Communicable, neonatal and nutrition causes rose
D: Non-communicable causes decreased and Communicable, neonatal and nutrition causes decreased

As shown in Figure 1, the two items have different difficulty levels reflecting their respective complexity level. The difficulty index provided by ExamSoft® shows how easy the item is and therefore complexity is measured as the complement of the provided index up to a total of 1 or 100%. As Figure 1 indicates, the first item, at a recall level, has a complexity level of 20% (easiness of 0.8 or 80%) while the second item, at an understanding level, has a complexity level of 54%. However, ExamSoft® also provides a discrimination index for both the correct choice and each distractor that allows the instructor to verify the quality of the question. Both questions have an acceptable discrimination index (0.30 and respectively 0.39) while the distractors have smaller, negative discrimination indices (Figure 1).

The item analysis data also offers insights on item situations that are not falling within the relatively standard situations described above. For example, consider two additional questions from this exam.

It is important for a heath professional to consider culture because:
A. Patients prefer that approach
B. It is a very "current" approach to medicine
✓C. Cultural beliefs regarding illnesses vary greatly
D. They really should not consider culture

The pharmacoeconomic evaluation approach in which benefits and costs are simply listed (but not evaluated) is considered the:

A. Cost-minimization analysis
✓B. Cost-consequence analysis
C. Cost-benefit analysis
D. Cost-effectiveness analysis
E. Cost-utility analysis

Item analysis output for these two questions is presented in Figure 2.

FIGURE 2
ITEM ANALYSIS INFORMATION FOR MASTERY AND VERY DIFFICULT MULTIPLE-CHOICE EXAM ITEMS ADMINISTERED IN AN "INTRODUCTION TO U.S. HEALTH CARE SYSTEMS" COURSE.

Question #	Correct Responses			Disc. Index	Point Biserial	Correct Answer	Response Frequ				
	Diff(p)	Upper	Lower				A	B	C	D	E
Description: q09											
4	0.99	100.00%	95.65%	0.04	0.11	C	1	0	*84	0	-
Questi ID / Rev: 7548 / 2				-	-	% Selected	1.18	0.00	98.82	0.00	-
-	-	-	-	-		Point Biserial (rpb)	-0.11	0.00	0.11	0.00	-
-	-	-	-	-		Disc. Index	-0.04	0.00	0.04	0.00	-
-	-	-	-	-	-	Upper 27%	0.00	0.00	1.00	0.00	-
-	-	-	-	-	-	Lower 27%	0.04	0.00	0.96	0.00	-

Q: It is important for a health professional to consider culture because:
A: patients' prefer that approach.
B: it is a very "current" approach to medicine.
* **C:** cultural beliefs regarding illness vary greatly.
D: they really should not consider culture.

Description: Research											
12	0.15	26.09%	0.00%	0.26	0.35	B	7	*13	45	5	15
Questi V Rev: 281 / 2				-	-	% Selected	8.24	15.29	52.94	5.88	17.65
-	-	-	-	-		Point Biserial (rpb)	-0.13	0.35	0.11	-0.11	-0.31
-	-	-	-	-		Disc. Index	-0.09	0.26	0.17	-0.04	-0.30
-	-	-	-	-	-	Upper 27%	0.04	0.26	0.65	0.00	0.04
-	-	-	-	-	-	Lower 27%	0.13	0.00	0.48	0.04	0.35

Q: The pharmacoeconomic evaluation approach in which benefits and costs are simply listed (but not evaluated) is considered the:
A: cost-minimization analysis
* **B:** cost-consequence analysis
C: cost-benefit analysis
D: cost-effectiveness analysis
E: cost-utility analysis

The first question addresses the definition and importance of culture in patient care. This question is complex because it requires both an understanding of the definitions of cultural competence and cultural awareness, as well as students' recognition of how a greater understanding of a patient's culture facilitates a more meaningful interaction with a patient. Response *"A"* indicates a very shallow understanding of these complexities, as the response focuses solely on patient satisfaction. Option *"B"* is similarly shallow as it motivates the importance of understanding patient cultures on political agendas within the clinician's profession.

Option *"D"* indicates a response that shows the lowest level of understanding of the important of cultural awareness and/or competence. Option *"C"* is correct, and respondents who choose this option display the deepest levels of understanding. These students not only grasp the complex nature of the link between culture and patient beliefs, but also how those patient beliefs shape the interaction between the patient and the clinician, and by extension the quality of care provided.

Because PharmD program offers a professional degree, the grading scheme is a mastery one, requiring a minimum of 70% as passing grade. Therefore, often multiple-choice items falling under the

mastery assessments will generate discriminant indices that are very low, as shown for the question in the top panel of Figure 2. Since the instructor made cultural competence an important part of the instructional process, this question was a mastery one and the discriminant index is not an indicator of this question's quality.

The second question addresses the basic modes of economic analysis used to evaluate clinical interventions. This question requires relatively low levels of learning as it requires students to know basic definitions. However, learning these definitions also requires an understanding of how the benefits of specific clinical interventions are characterized and combined with costs to create an overall evaluation of that intervention, relative to another established, comparable clinical intervention. Option *"A"* requires no specific evaluation of relative benefits, as they are equal across the two interventions. Only the costs of the two interventions are compared. Options *"C"*, *"D"*, and *"E"* measure benefits in currency, clinical values, and patient subjective evaluations, respectively. However, in all instances, implicitly understanding these definitions requires knowledge that benefits and costs *can* be evaluated in a relative and quantitative sense. Only option *"B"*'s definition allows a listing of costs and benefits in non-comparable terms, and are as such not evaluated in a comparative context.

For the more challenging questions that yield student response distributions similar to those identified in the lower panel of Figure 2, the instructor must use her/his judgement to determine whether the question:

(a) Is simply assessing difficult material and deserving of greater reinforcement in a later part of the course, or

(b) It reflects a strong misconception that necessitates additional in-depth instructional interventions to correct.

While the discriminant index for this difficult question is close to an acceptable level of 0.30, the fact that only 15% of students selected the correct answer warrants a more close analysis of this item. A quick scan of answer distribution shows that 53% of the students selected the distractor that has the highest probability to be the most often discussed topic in this part of the course, the cost-benefit analysis. This observation points toward a potential explanation for this questions difficulty, a superficial read of the stem by students who missed the key element in the question stem, the fact that the two elements are not evaluated "(but not evaluated)". Regardless, this finding suggests the need for a more in depth analysis of the instructional process that supports the learning process associated with this assessment item and possible inclusion, for example, of some real-time feedback through clicker questions to help students better differentiate among various pharmacoeconomic evaluation strategies presented in the course.

In addition to the punctual, exam-based item analysis, the instructor can compare assessment items across multiple semesters to support assessment decisions such as item improvement or dismissal. Figure 3 shows an item improvement situation that involved a restructuring of the item distractors. The initial question was written as follows:

In Africa, primary health care clinics have which of the following advantages?

A. They allow physicians to see many patients in a single day
B. The can provide follow-up care for patients discharged from hospitals
C. They provide free medications to patients seen at the clinic
D. A and B
✓E. B and C

This question, focused on global health, addresses the importance of primary care in developing countries. The correct answer to the question requires students to understand that patients in under-developed countries cannot afford (expensive) medications and, that those patients that have acute medication conditions lack access to follow up care once discharged from a hospital. Primary care clinics fill this gap in access.

As shown in Figure 3, the instructor decided to drop two of the initial distractors and add a new one with the goal of guiding students toward the focus of this item, the identification of the target context-based advantages discussed in the lecture.

The question now reads as follows:

In Africa, primary health care clinics have which of the following advantages?

A. The can provide follow-up care for patients discharged from hospitals
B. They provide free medications to patients seen at the clinic
✓C. A and B
D. None of the above

FIGURE 3
MULTI-SEMESTER ANALYSIS INFORMATION FOR AN INDIVIDUAL EXAM ITEM: (A) INITIAL DEPLOYMENT AND ANALYSIS; (B) DEPLOYMENT AND ANALYSIS OF THE REVISED ASSESSMENT ITEM.

Question #	Correct Responses			Disc. Index	Point Biserial	Correct Answer	Response Frequ				
	Diff(p)	Upper	Lower				A	B	C	D	E
Description: Global HealthyReform											
34	0.52	62.50%	48.28%	0.14	0.13	E	2	2	9	26	*42
Question Rev: 256 / 4				-	-	% Selected	2.47	2.47	11.11	32.10	51.85
	-	-	-	-		Point Biserial (rpb)	-0.10	0.04	0.07	-0.16	0.13
-	-	-	-	-	-	Disc. Index	-0.03	0.01	0.06	-0.18	0.14
-	-	-	-	-	-	Upper 27%	0.00	0.04	0.17	0.17	0.63
-	-	-	-	-	-	Lower 27%	0.03	0.03	0.10	0.34	0.48

Q: In Africa, Primary Health Care clinics have which of the following advantages?
A: They allow physicians to see many patients in a single day
B: They can provide follow-up care for patients discharged from hospitals
C: They provide free medications to patients seen at the clinic
D: a and b
* E: b and c

(a) Spring Semester

Question #	Correct Responses			Disc. Index	Point Biserial	Correct Answer	Response Fre				
	Diff(p)	Upper	Lower				A	B	C	D	E
33	0.71	86.96%	56.52%	0.30	0.29	C	5	12	*60	8	-
Question Rev: 256 / 5				-		% Selected	5.88	14.12	70.59	9.41	-
	-	-	-	-		Point Biserial (rpb)	0.11	-0.26	0.29	-0.23	-
-	-	-	-	-	-	Disc. Index	0.04	-0.17	0.30	-0.17	-
-	-	-	-	-	-	Upper 27%	0.04	0.09	0.87	0.00	-
-	-	-	-	-	-	Lower 27%	0.00	0.26	0.57	0.17	-

Q: In Africa, Primary Health Care clinics have which of the following advantages?
A: They can provide follow-up care for patients discharged from hospitals
B: They provide free medications to patients seen at the clinic
* C: a and b
D: none of the above

(b) Fall Semester

This change brought the discrimination index to an acceptable level and dropped the difficulty of the question to an acceptable mastery level for this examination.

Finally, Figures 4a and 4b shows how comparing the item analysis data across semesters can help identify low-performing assessment items that might need discharged unless some major changes could be planned in the instructional process to make sure the learning outcome targeted by those items have appropriate scaffolds to support student learning.

The first question (Figure 4a) is a question that requires basic remembering of the goals of the Millennium Development initiative.

Which of the following is NOT one of the eight Millennium Development Goals:

A. Eradicate extreme poverty and hunger

✓B. Expand hospital services to achieve equality

C. Promote gender quality and empower women

The question analyzed in Figure 4b deals with domestic health disparities, rather than international disparities:

Which of the following groups had the worst access to care compared to whites per the 2011 National Healthcare Quality and Disparities Report?

A. Black/African American

B. Asians

✓C. Hispanics

As shown in Figure 4b, in its first deployment this item proved to be a rather difficult one (difficulty of about 58%) while the discrimination index had quite a low value below the accepted borderline. For the second deployment, the difficulty increased to 62% but the discrimination index became close to the accepted borderline values accepted for difficult items.

One of the possible reasons for this item's student-specific behavior is that it references to a very specific report that, while covered in the lecture, might not have captured students' full attention when they prepared for the exam. This is especially poignant considering that data in these types of reports usually becomes quite rapidly obsolete. Another possible reason for these findings (whose resolution is likely beyond the scope of the current study) might be a cultural stereotype in the context and wording of the question and its structure that leads students to select option *"A"*. Considering these possible rationale, the instructor has the option to restructure this question or replace it with a different one associated with the same target-learning outcome.

Assessment Evaluation beyond Individual Items

The analysis of individual items described previously is a critical step to ensure the effectiveness of a given assessment. However, the instructor can also use ExamSoft® to generate detailed reports by various learning objectives (e.g. ability-based outcomes) and learning taxonomical levels (e.g., Bloom's taxonomy); these analyses can be conducted for both individual exams, and longitudinally across chosen exams or semesters.

Figure 5 exemplifies a longitudinal report for the "Introduction to U.S. Health Care Systems" course that ties questions to their specific level of Bloom's taxonomical levels. To provide a more meaningful analysis, Figure 5 was constructed by generating an ExamSoft® report for the Exam 3 administered during the spring 2015 semester and respectively during the fall 2016 semester. The results from the analysis of longitudinal data can provide the instructor with a critical perspective on the level of coverage of various critical elements throughout the entire instructional process and at the same time can provide critical input for accreditation purposes. In this case, the results suggest an opportunity for student growth in applying knowledge learned in the course. Interestingly, students in the course have fewer issues with analysis (a higher order skill on Bloom's Taxonomy) than they do with the lower order skills of remembering, understanding, and applying course content.

FIGURE 4
ITEM ANALYSIS INFORMATION FOR LOW-PERFORMING ASESSMENT ITEMS
(A) VERY DIFFICULT MULTIPLE-CHOICE EXAM ITEMS

Question #	Correct Responses			Disc. Index	Point Biserial	Correct Answer	Response Fr				
	Diff(p)	Upper	Lower				A	B	C	D	E
23	0.26	41.67%	10.34%	0.31	0.26	B	14	*21	46	-	-
Question	Rev: 267 / 3			-	-	% Selected	17.28	25.93	56.79	-	-
	-	-	-	-		Point Biserial (rpb)	0.06	0.26	-0.27	-	-
-	-	-	-	-	-	Disc. Index	0.04	0.31	-0.35	-	-
-	-	-	-	-	-	Upper 27%	0.21	0.42	0.38	-	-
-	-	-	-	-	-	Lower 27%	0.17	0.10	0.72	-	-

Q: Which of the following is NOT one of the eight Millenium Development Goals:
A: Eradicate extreme poverty and hunger
* **B:** Expand hospital services to achieve equality
C: Promote gender equality and empower women

(a) Spring Semester

Question #	Correct Responses			Disc. Index	Point Biserial	Correct Answer	Response Frequ				
	Diff(p)	Upper	Lower				A	B	C	D	E
Description: Research											
22	0.25	21.74%	26.09%	-0.04	0.00	B	7	*21	57	-	-
Question	Rev: 267 / 3			-	-	% Selected	8.24	24.71	67.06	-	-
	-	-	-	-		Point Biserial (rpb)	0.09	0.00	-0.06	-	-
-	-	-	-	-	-	Disc. Index	0.04	-0.04	0.00	-	-
-	-	-	-	-	-	Upper 27%	0.09	0.22	0.70	-	-
-	-	-	-	-	-	Lower 27%	0.04	0.26	0.70	-	-

Q: Which of the following is NOT one of the eight Millenium Development Goals:
A: Eradicate extreme poverty and hunger
* **B:** Expand hospital services to achieve equality
C: Promote gender equality and empower women

(b) Fall Semester

(B) DIFFICULT MULTIPLE-CHOICE EXAM ITEMS

Question #	Correct Responses			Disc. Index	Point Biserial	Correct Answer	Response Frequ				
	Diff(p)	Upper	Lower				A	B	C	D	E
Description: Developments											
5	0.42	45.83%	27.59%	0.18	0.11	C	45	2	*34	-	-
Question	Rev: 249 / 3			-	-	% Selected	55.56	2.47	41.98	-	-
	-	-	-	-		Point Biserial (rpb)	-0.15	0.11	0.11	-	-
-	-	-	-	-	-	Disc. Index	-0.22	0.04	0.18	-	-
-	-	-	-	-	-	Upper 27%	0.50	0.04	0.46	-	-
-	-	-	-	-	-	Lower 27%	0.72	0.00	0.28	-	-

Q: Which of the following groups had the worst access to care compared to whites per the 2011 National Healthcare Quality and Disparities Report?
A: Black/African American
B: Asians
* **C:** Hispanics

(a) Spring Semester

Question #	Correct Responses			Disc. Index	Point Biserial	Correct Answer	Response Frequ				
	Diff(p)	Upper	Lower				A	B	C	D	E
Description: Developments											
38	0.38	65.22%	30.43%	0.35	0.28	C	47	6	*32	-	-
Question	Rev: 249 / 4			-	-	% Selected	55.29	7.06	37.65	-	-
	-	-	-	-		Point Biserial (rpb)	-0.17	-0.20	0.28	-	-
-	-	-	-	-	-	Disc. Index	-0.22	-0.13	0.35	-	-
-	-	-	-	-	-	Upper 27%	0.35	0.00	0.65	-	-
-	-	-	-	-	-	Lower 27%	0.57	0.13	0.30	-	-

Q: Which of the following groups had the worst access to care compared to whites per the 2011 National Healthcare Quality and Disparities Report?
A: Black/African Americans
B: Asians
* **C:** Hispanics

(b) Fall Semester

While the analysis presented in Figure 5 reflects the results from only two successive exams, the instructors can produce equivalent data for all multiple-choice assessments administered through ExamSoft® across time providing a useful snapshot of item complexity distribution.

FIGURE 5
LONGITUDINAL ANALYSIS BY BLOOM'S TAXONOMICAL LEVELS FOR MULTIPLE-CHOICE ITEMS ADMINISTERED IN THE "INTRODUCTION TO U.S. HEALTH CARE SYSTEMS" COURSE DURING THE SPRING 2015 AND FALL 2016 SEMESTERS

	BL1 – Remember	BL2 - Understand	BL3 - Apply	BL4 - Analyze
# Assessments	2	2	1	2
# Items	4	44	5	6
Group Average	76.96%	77.43%	66.82%	83.33%
min	25.00%	57.14%	40.00%	50.00%
max	100.00%	94.29%	100.00%	100.00%
median	75.00%	77.50%	60.00%	83.33%
St Dev	19.27%	7.66%	16.78%	15.46%

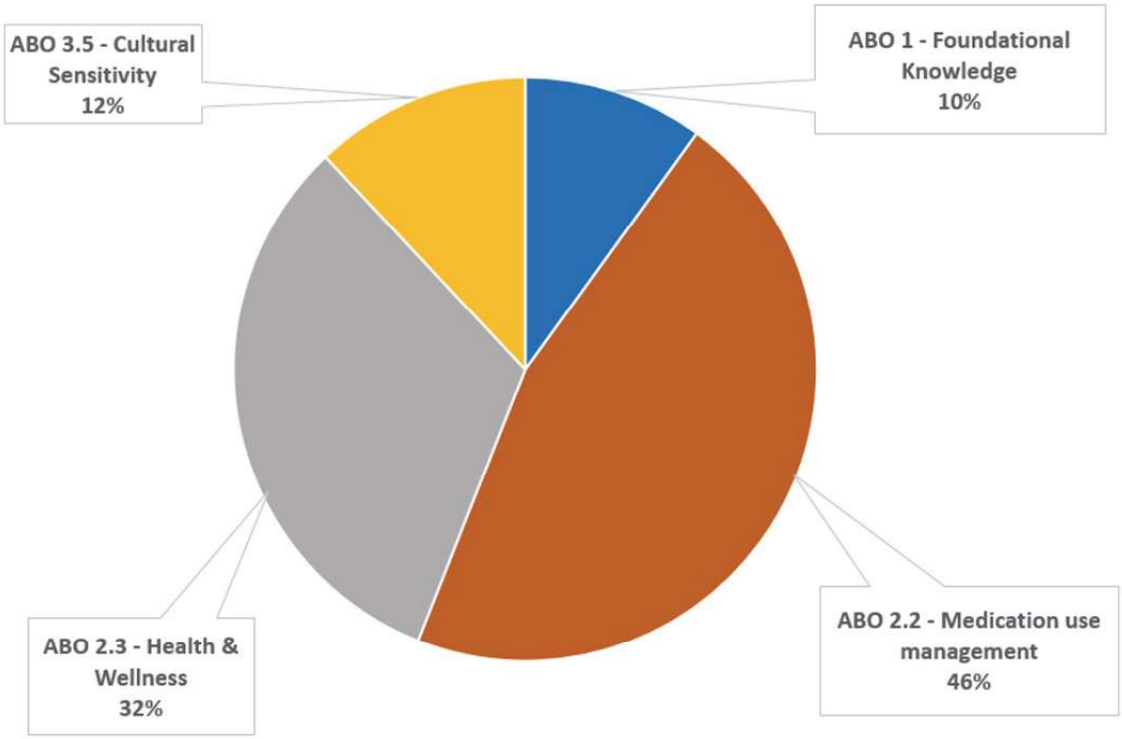
ExamSoft® also allows the instructor to export the data in formats that can produce a helpful data visualization associated with the instructional process reflected in the exam (or the semester). For example, Figure 6 shows the coverage of the major ability based outcomes (ABOs) for the entire "Introduction to U.S. Health Care Systems" course during the spring 2015 semester.

This diagram can help the instructor analyze to what degree the administered assessments reflect the desired coverage of target outcomes as expected for the level and type of course. Similar visuals derived from other important tracked categories such as taxonomy level or course objective can also help the instructor to compare what and how much was assessed with what and how much content was covered during the instructional process. Perhaps more importantly, as faculty use the information generated by students' incorrect responses more effectively to improve course content and delivery, the information in Figures 5 and 6 can provide a more global context in which to assess the effectiveness of those improvements.

DISCUSSION

In this exploratory case study, we illustrated how an analysis of students' multiple-choice responses, both correct and incorrect ones, provided additional information useful for instructors in improving future student performance and implicit supporting their learning process. Our results indicate that incorrect responses were more informative in identifying and addressing curricular gaps than the correct responses. The use of an exam administration software such as ExamSoft® will facilitate this process by providing a simple means to collate and analyze exam results.

FIGURE 6
DISTRIBUTION OF ASSESSMENT ITEMS BY ACTUAL ABILITY BASED OUTCOMES (ABO) FOR EXAM 3 MULTIPLE-CHOICE ITEMS ADMINISTERED IN THE "INTRODUCTION TO U.S. HEALTH CARE SYSTEMS" COURSE DURING THE SPRING 2015 AND FALL 2016 SEMESTERS.



Taken collectively, intentional, evidence-based instructional design and assessment yields improved student learning. ExamSoft® also empowers faculty and administrators to identify and track over time evidence-based teaching improvements and, in doing so, to support the curricular evaluation and accreditation processes.

Since the improvement of assessment presented in this exploratory case study can span from the individual instructor, to department and organization level, the use of ExamSoft® creates a strong basis for moving assessment and curricular improvement from a punctual one-time analysis to an innovative all-inclusive process driven by successful mining of big assessment datasets.

Limitations

As this paper provides an exploratory case study, its analysis is by no means exhaustive. For simplicity, the case study focused on only specific exam questions. In practice, instructors would replicate this process for all major assessment items in a course. One limitation that is specific to a successful implementation of this process is the ability to secure a strong administrative support from the onset of this activity because it requires a significant time commitment on the part of the instructor. First, the instructor needs support with training and upload of the exams from their typical Microsoft Word® format to the web-based format required by the used software.

At our institution, this commitment consists of an assistant professor/ instructional designer that coordinate these activities and a graduate service assistant that uploads the exam questions in ExamSoft®.

In addition, for the instructor the commitment lies with the thoughtful design and redesign of the questions and possible responses. If an instructor is not intentional in constructing the incorrect question

to characterize specific types of learning gaps, students who select these incorrect responses will not signal any useful information about their learning gaps.

Moreover, to tie this additional information to curricular outcomes and/or conduct longitudinal assessments, the intentional question design must be implemented consistently throughout one's course. The authors argue that this effort is a reasonable investment, given the analysis capabilities of the software.

CONCLUSIONS

This study presents the early positive implications of the integrations of an exam software, ExamSoft®, into the assessment process with focus on the instructor's ability to enhance significantly the effectiveness of multiple-choice assessment items with benefits for students' learning. The longitudinal analysis capability of this examination software expands the potential implications of these improvements from exam to curricular mapping and accreditation support levels.

We intend to further pursue this first study into the analysis of the implications of changes in the quality of assessment items across semesters with a major focus on students' learning performance. In addition, we intend to track those assessment items that reflect major misconceptions in students' learning, develop instructional interventions to address them and measure the impact of these interventions across semesters.

Using the big data analysis capabilities of ExamSoft® we intend to use the results of these further analyses to expand the data-driven decisions from instructor-level to curricular level for the PharmD program at our institution.

REFERENCES

ACPE. (2015). Accreditation Standards and Key Elements for the Professional Program in Pharmacy Leading to the Doctor of Pharmacy Degree. Retrieved from https://www.acpe-accredit.org/pdf/Standards2016FINAL.pdf

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., . . . Wittroch, M. C. (2001). *A taxonomy for learning, teaching, and assessing. A revision of Bloom's taxonomy of educational objectives*. New York: NY: Addison Wesley Longman, Inc.

Bereby-Meyer, Y., Meyer , J., & Flascher, O. M. (2002). Prospect Theory Analysis of Guessing in Multiple Choice Tests. *Journal of Behavioral Decision Making*, 15, 313-327. doi:10.1002/bdm.417

Burton, R. F. (2002). Misinformation, partial knowledge and guessing in true/false tests. *Medical Education*, 36, 805-811.

Downing, S. M. (2002a). Threats to validity of locally developed multiple-choice tests in medical education: Construct-irrelevant variance and construct underrepresentation. *Advances in Health Sciences Education*, 7, 235-341.

Downing, S. M. (2002b). Construct-irrelevant variance and flawed test questions: Do multiple-choice item-writing principles make any difference? . *Academic Medicine*, 77(10), S103-S104.

Downing, S. M. (2003). Commentary: Guessing on selected-response examinations. *Medical Education*, 37, 670-671.

Morrison, G. R., Ross, S. M., Kalman, H. K., & Kemp, J. E. (2011). *Designing Effective Instruction* (Sixth ed.). Hoboken, NJ: John Wiley & Sons, Inc.