

Using Multiple Choice Questions Written at Various Bloom's Taxonomy Levels to Evaluate Student Performance across a Therapeutics Sequence

Amy M. Tiemeier, Pharm.D., BCPS; Zachary A. Stacy, Pharm.D., BCPS; and John M. Burke, Pharm.D., FCCP, BCPS
St. Louis College of Pharmacy, Saint Louis, Missouri

Keywords: Multiple Choice Question; Bloom's Taxonomy; Therapeutics; Pharmacy

Abstract

Objective: To evaluate the results of a prospectively developed plan for using multiple choice questions (MCQs) developed at defined Bloom's levels to assess student performance across a Therapeutics sequence.

Methods: Faculty were prospectively instructed to prepare a specific number of MCQs for exams in a Therapeutics sequence. Questions were distributed into one of three cognitive levels based on a modified Bloom's taxonomy, including recall, application, and analysis. Student performance on MCQs was compared between and within each Bloom's level throughout the Therapeutics sequence. In addition, correlations between MCQ performance and case performance were assessed.

Results: A total of 168 pharmacy students were prospectively followed in a Therapeutics sequence over two years. The overall average MCQ score on 10 exams was 68.8%. A significant difference in student performance was observed between recall, application, and analysis domain averages (73.1%, 70.2% and 60.1%; $p < 0.001$). Student performance within each Bloom's level across the three courses was significantly different for recall ($p < 0.001$), application ($p < 0.001$), and analysis ($p < 0.001$) MCQs. A significant correlation was observed between the recall domain and the case (0.67; $p < 0.01$), application domain and the case (0.62; $p < 0.01$), and analysis domain and the case (0.64; $p < 0.01$).

Conclusions: As students progress through the curriculum, faculty may need to find ways to promote recall knowledge for more advanced topics while continuing to develop their ability to apply and analyze information. Exams with well-designed MCQs that prospectively target various cognitive levels can facilitate assessment of student performance.

Background

In pharmacy schools across the country, faculty strive to ensure that students are achieving course outcomes. The Accreditation Council for Pharmacy Education (ACPE) provides standards pertaining to student learning and assessment of learning. Standard 11 (Teaching and Learning Methods) states that we must foster the "development and maturation of critical thinking and problem-solving skills" in our students.¹ Standard 15 (Assessment and Evaluation of Student Learning and Curricular Effectiveness) mandates that we assess this higher level learning with methods that "employ a variety of valid and reliable measures systematically and sequentially throughout the professional degree program."¹ The challenge is to appropriately develop courses so that assessment methods adequately measure the higher level abilities as well as knowledge. Faculty should develop a process by which a variety of cognitive levels are evaluated during testing.²

One method for characterizing levels of learning was defined in the 1950's by educational psychologist Benjamin Bloom.

In his work, Bloom proposed six hierarchical and cumulative levels of student learning (Figure 1).³ Lower levels of learning focus on recall of information while middle levels require application of knowledge. Higher thinking levels involve the deconstruction and construction of concepts. Applying this model to testing can be helpful in evaluating student performance at various cognitive levels.⁴

An assessment plan must be constructed deliberately if student performance is to be measured at multiple cognitive levels. The use of open-ended questions may be helpful in assessing a student's application or analysis abilities. However, it need not be the only method. If appropriately constructed, multiple choice questions (MCQs) may also assess performance at higher levels, but specific efforts must be taken to ensure the targeting of various cognitive levels. Similarly, if student abilities are to develop across a sequence of courses, coordination of an assessment plan is necessary. This is particularly important in a team-taught course where faculty may develop questions without consideration of how other faculty may be assessing students.

The advantages and disadvantages of MCQs should be weighed when deciding their role in a course assessment plan. Perhaps the most common reason for their use is quick and objective grading.⁵ Electronic course management

Corresponding Author: John M. Burke, Pharm.D., FCCP, BCPS
4588 Parkview Place, Saint Louis, Missouri 63110-1088
Phone: 314-446-8506, Fax: 314-446-8500,
Email: jack.burke@stlcpop.edu

systems also facilitate accurate and timely recording of scores. In addition, the same question can be re-administered to reassess or compare performance of individuals or groups. While administration of MCQs may be simple, development of high quality questions requires a significant time investment, with one author estimating 1 hour per question.⁵ Because there are a limited number of choices, a correct answer does not necessarily assure student understanding. Item analysis, however, can provide some insight into the ability of a question to discriminate student performance.⁶ Additionally, a common perception is that MCQs can only test lower levels of thinking while fill-in-the-blank, short answer, or essay questions are necessary to assess higher levels of thinking. However, several publications have documented that MCQs can successfully be used to test at higher cognitive levels.⁷⁻⁹ Moreover, MCQs have been shown to be comparable to essay questions when used for higher order testing.⁷ Therefore, MCQs may be able to play a role in assessing students' performance at a variety of levels, if deliberately incorporated into a course assessment plan.

This prospective cohort study was conducted at a private six-year pharmacy school. The college enrolls approximately 1200 students and employs 45 full-time pharmacy practice faculty. The patient care sequence begins in the spring semester of the first professional year (year 3) with Pathophysiology (5-credit hour course). This is followed by consecutive semesters of Therapeutics course work (4-credit hours each) through the second and third professional years. Specific lecture topics assigned to each Therapeutics course can be found in Figure 2. Each course includes lectures provided by a team of faculty to approximately 175 students and weekly case discussions with groups of 25-30 students. While each course uses examinations composed of MCQs and application-based cases, the distribution of MCQs and case points varies based on course and exam. Traditionally, the cognitive level of a MCQ was independently decided by the lecturer resulting in random distribution of questions at unspecified levels of difficulty on each exam. This study was designed to evaluate the results of a prospectively developed plan using MCQs developed at defined Bloom's levels to assess student performance across a Therapeutics sequence.

Methods

Faculty were prospectively instructed to prepare a specific number of MCQs for exams in a Therapeutic sequence. Questions were distributed into one of three cognitive levels based on a modified Bloom's taxonomy, including recall, application, and analysis (Figure 1). "Recall" questions included the knowledge and comprehension cognitive levels. A recall question only required knowledge or basic understanding of a fact to answer. These exam questions may

ask the student to identify or define factual information. "Application" questions required students to apply knowledge, whereas "analysis" questions required students to analyze, synthesize, or evaluate information. An application question required applying knowledge, perhaps including a calculation or interpretation of information. In addition, exam questions written for the application level may ask a student to solve a problem or classify information. An analysis question required interpretation of numerous facts to answer a multistep problem. The student must evaluate the problem and perform the required, implicit steps to determine a solution. Each examination was standardized to contain a ratio of MCQs distributed as 40% recall, 40% application, and 20% analysis. The number and distribution of MCQs were evenly distributed among topics based on the lecture hours per topic. Each exam covered between 11 and 15 hours of lecture time.

Investigators evaluated each MCQ and its proposed Bloom's classification. If the three primary investigators agreed that the MCQ was written at the desired level, the question was accepted for use on the exam. An independent faculty member was consulted on occasion for assistance in classifying the question. When a MCQ was not written at the desired level, investigators worked with faculty to re-write the MCQ at the desired level before inclusion on the exam. The percentage of questions initially written at the requested Bloom's level was recorded. Question writers were encouraged to use a submission form to facilitate the peer review process which identified the question writer, lecture topic, question, answer, and Bloom's level (Figure 3).

Examinations were composed of two sections including a case and MCQs. The case component included a ½-1 page patient case followed by a series of fill-in-the-blank questions designed to challenge students with open-ended questions. The case comprised 20-25% of the total exam points. The MCQ component was completed using traditional scantron technology (OpScan-3, River Falls, WI), while the paper-based short-answer case was graded by hand. Each MCQ contained a question stem followed by one correct answer and three plausible, but incorrect choices. Negatively-phrased and K-type MCQs were not permitted.

The study included students enrolled in Therapeutics courses in the second and third professional years of the curriculum. Students repeating a course or those students who did not progress to the subsequent course were excluded, so that each student could serve as his/her own control throughout the study.

The primary outcome was the comparison of student performance on MCQs between each Bloom's level throughout the Therapeutics sequence. One secondary outcome evaluated student performance across the Therapeutics curriculum within each Bloom's level. This was undertaken to determine if there was a change in performance at each Bloom's level as a student progresses through the curriculum. A comparison of student performance on MCQs to their performance on written examination cases was undertaken to compare these two assessment methods. It was postulated that performance at higher Bloom's levels might correlate with performance on the written cases. A detailed examination key was provided for each faculty member grading the written cases to facilitate consistency in grading.

Student performance was measured and reported as the average percent correct in each Bloom's level. An ANOVA test was used for the primary comparison of student performance in each Bloom's level. Correlations were analyzed using the Pearson's Correlation Coefficient. Exam reliability was calculated using the Cronbach-Alpha formula. Cronbach's alpha ranges from 0 to 1.00, with values close to 1.00 indicating high consistency. Statistical analyses were performed using SPSS, version 16.0 and Microsoft Excel 2007. This research has been reviewed and approved for exempt status by the St. Louis College of Pharmacy Investigational Review Board.

Results

This prospective evaluation included 168 pharmacy students. A total of 16 students were excluded either due to delay in progression ($n=12$) or withdrawal from the college ($n=4$) during the study. Ten exams were administered throughout the Therapeutics sequence which included a total of 116 recall (40.1%), 114 application (39.5%), and 59 analysis (20.4%) MCQs. Of the 289 MCQs in the three courses, 82.7% of the questions were initially written at the requested Bloom's level. A similar number of MCQs were written correctly at the prescribed level in each of the three courses (86.0%, 80.0%, 81.0%; $p>0.05$). The exam reliability score for the MCQ section ranged from 0.402 to 0.611.

The overall MCQ and case average scores on the 10 exams were 68.8% and 62.3%, respectively. A significant difference in student performance was observed between recall, application, and analysis domain averages (73.1%, 70.2% and 60.1%; $p<0.001$) (Figure 4). In addition, when analyzing student performance within each Bloom's level across the three Therapeutics courses, a significant difference was observed for recall (74.8%, 71.3%, 69.0%, $p<0.001$), application (69.4%, 64.7%, 75.5%, $p<0.001$), and analysis

MCQs (54.7%, 59.8%, 66.8%, $p<0.001$) (Figure 5). While performance on analysis questions progressively improved, performance on recall questions worsened.

A significant correlation was observed between the recall domain and the case (0.67; $p<0.01$), application domain and the case (0.62; $p<0.01$), and analysis domain and the case (0.64; $p<0.01$). Figure 6 compares each student's average scores on their recall MCQs (x axis) to their average scores on the cases (y axis). The scatter plots for application and analysis MCQs compared to cases scores were similar. This suggests that the use of MCQs at each of the Bloom's levels would provide a similar assessment of student performance as the use of cases.

Discussion

In this study, we examined how students perform in three modified categories based on Bloom's taxonomy. A distribution of 40% recall, 40% application, and 20% analysis was chosen based on a previous experience in a professional course. Wong and colleagues studied the effectiveness of an exam blueprint based on Bloom's taxonomy in a pharmacy curriculum within an oncology block. Varying ratios of lower and higher level learning MCQs were assessed across three exams (75:25, 50:50, 40:60). Performance on lower level questions did not significantly change (89%, 74%, 84%); however, aptitude at higher level questions did improve (60%, 77%, 83%).¹⁰ Varying the distribution of MCQs across cognitive levels may be appropriate if the content is narrowly focused or with repeat assessments over time. We chose to use a fixed ratio as the content varied, often becoming more difficult, as the students progressed through the curriculum. The use of three cognitive levels was chosen for ease of implementation in the Therapeutics sequence. With minor differences existing between some Bloom's levels, our modified levels categorized MCQs in a way that highlights the three ways to use information for clinical problem solving.

Using cases with open-ended questions does not necessarily assess performance at higher cognitive levels. Our investigation demonstrated a significant correlation between performance on written cases and MCQs at each of the three Bloom's levels. Further evaluation of the cognitive level of case questions may be necessary to ensure desired outcomes are achieved. Just as all MCQs do not necessarily test at the knowledge level, short answer case questions do not always test students at the application or analysis level.¹¹

A longitudinal investigation was performed to determine if student performance improved across various cognitive levels. While students improved at the analysis level, their performance on recall level questions decreased. The

worsening of lower level cognitive scores may not be indicative of the students' diminished ability to recall and apply information, rather inexperience with new and more advanced topic material. This suggests that critical thinking abilities are context dependent despite the students' experiences in earlier courses. As students progress through the curriculum, we may need to find ways to promote recall knowledge for more advanced topics while continuing to develop their ability to apply and analyze information. In addition, student performance at all cognitive levels may improve if courses are re-designed with practice opportunities that promote development of these higher level abilities.

Several limitations in methodology may affect the extrapolation of these data. With the absence of a control group, we cannot definitively conclude that the correlations found in this study would extrapolate to other courses, students, or colleges. This investigation studied the same students throughout a sequence of Therapeutics courses. While exam reliability scores were collected for overall assessment of exam quality, the addition of an item discrimination analysis, such as the point biserial metric, should be considered in the future to ensure the consistency and quality of individual MCQs. Lastly, the investigators reached a consensus when analyzing the cognitive level of exam questions and did not objectively evaluate inter-rater reliability. Future studies should incorporate a technique to determine the consistency of agreement among raters using a kappa statistic.

Targeting MCQs at a specific Bloom's level was a new experience for many faculty. Nevertheless, over 80% of questions were initially written at the targeted cognitive level. Faculty had varying levels of experience and comfort with writing MCQs. Faculty development in this area may improve the ability and confidence of faculty in developing quality student assessments.¹²

Prospectively defining expectations for students may help in designing learning experiences and assessments. There should be congruency between learning objectives and the MCQs which assess achievement of the objective. In our study, the process of prospectively determining the level of MCQs was beneficial in refocusing the content and method of delivery. Faculty noted that this challenged them to think of new ways to assess student's abilities. The way material is taught may influence how a MCQ is classified. For example, while a question may appear to be written at the application level, it may only require students to recall information if it was taught in such a way that the application was done for the students. Focusing on the level of performance expected

of students may be helpful in designing optimal learning experiences.¹³

Multiple choice questions can be useful if Bloom's taxonomy is used to facilitate assessment of these abilities. Given the results shown in the scatterplots which showed a correlation between MCQ score to case score (e.g. Figure 6), it is feasible that each of the Bloom's levels evaluated in this study could be tested with MCQs. This makes it possible to use MCQs to test not only recall of knowledge, but the skills required to analyze and apply information. Hence, strategic use of MCQs can be valuable for assessments within a course, across courses, and may also be useful for milestone evaluations external to a specific course. This may be helpful in assessing student critical thinking and problem solving abilities that we hope to develop through our curriculum.

Summary

Well-designed MCQs can be a useful component of assessment in a Therapeutics sequence. Exams with MCQs that prospectively target various cognitive levels can facilitate this process. Faculty development may be necessary to optimize the use of MCQs as an assessment tool. Expanding this process to include other didactic and experiential courses may be helpful in assessing our entire pharmacy curriculum.

Acknowledgements

The investigators would like to acknowledge Theresa Prosser, Pharm.D. and Thomas Zlatic, Ph.D. for their expertise and assistance. We would also like to thank pharmacy practice faculty for their patience and cooperation with MCQ development.

References

1. Accreditation Council for Pharmacy Education. Accreditation standards and guidelines for the professional program in pharmacy leading to the doctor of pharmacy degree. Adopted January 15, 2006. Available at: <http://www.acpe-accredit.org>. Accessed June 2010.
2. Background Paper II: Entry-Level, Curricular Outcomes, Curricular Content and Educational Process; Commission to Implement Change in Pharmaceutical Education. *Am J Pharm Educ* 1993;57: 377-85.
3. Bloom B, Englehart M, Furst E, Hill W, Krathwohl D: Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York, Toronto: Longmans, Green; 1956.
4. Zlatic TD. Developing Thinking Abilities within Pharmacy Education: A Sourcebook. St. Louis College of Pharmacy and American Association of Colleges of Pharmacy, St. Louis. 1995.

5. Masters JC, Hulsmeyer BS, Pike ME, Leichty K, Miller MT, Verst AL. Assessment of multiple choice questions in selected test banks accompanying text books used in nursing education. *J Nursing Ed* 2001;40(1):25-31.
6. Morrison S, Free KW. Writing Multiple-Choice Test Items that Promote and Measure Critical Thinking. *J Nurs Ed* 2001; 40(1): 17-24.
7. Palmer EJ, Devitt PG. Assessment of Higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? *BMC Medical Education* 2007; 7: 49-55.
8. Palmer E, Devitt P: Constructing multiple choice questions as a method for learning. *Ann Acad Med Singap* 2006, 35:604-08.
9. Buckwalter JA, Schumacher R, Albright JP, Cooper RR. Use of an educational taxonomy for evaluation of cognitive performance. *J Med Educ* 1981, 56:115-21.
10. Wong SF, Quist R, Murray W. Effectiveness of examination blueprints to promote higher cognitive level performance. *Am J Pharm Educ* 2006;70(3):Article 65.
11. McDaniel TR. Designing essay questions for different levels of learning. *Improving College and University Teaching*. 1979;27(3):120-3.
12. A Guide to Preparing Multiple-Choice Items. New York, NY: Professional Examination Service, Inc.; c2005.
13. Schultheis NM. Writing cognitive educational objectives and multiple-choice test questions. *Am J Health Syst Pharm*. 1998; 55:2397-2401.

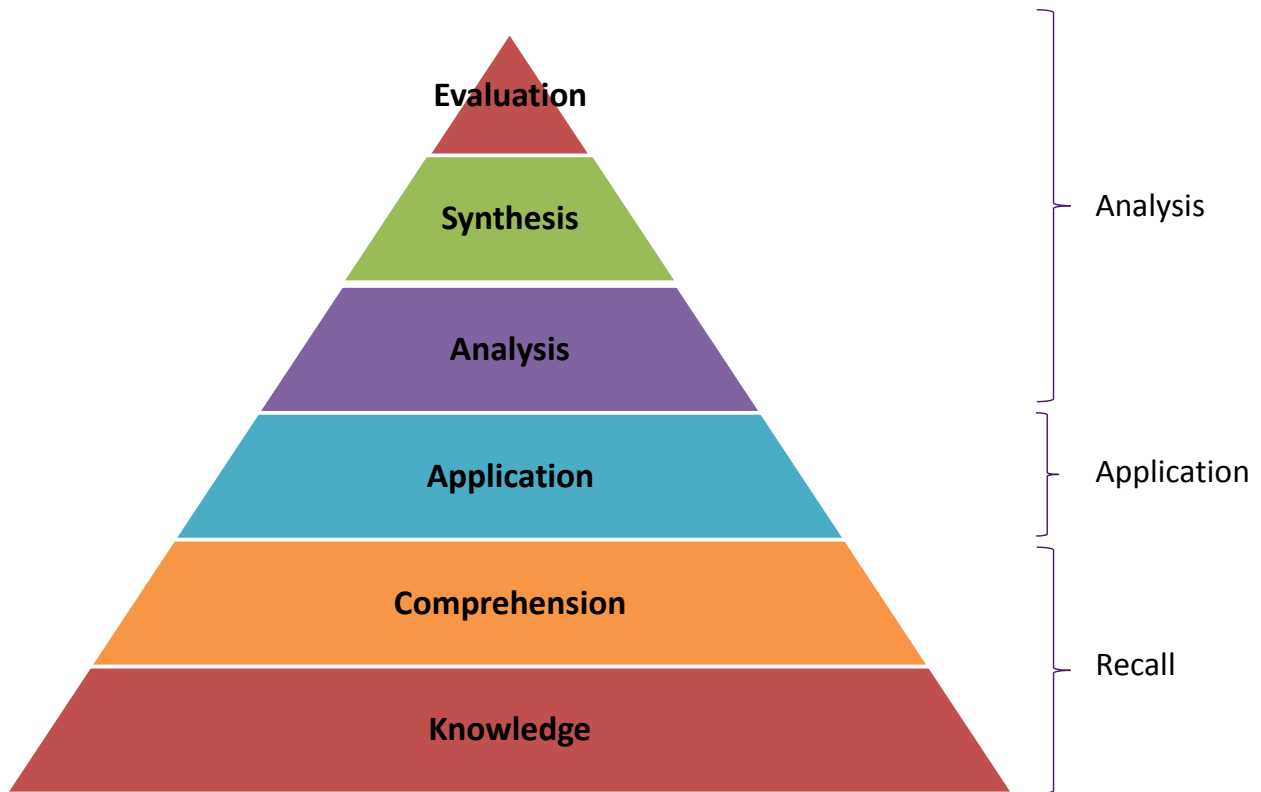


Figure 1: Bloom's Taxonomy is traditionally shown in a pyramid with lower cognitive levels at the base and higher cognitive levels at the apex.

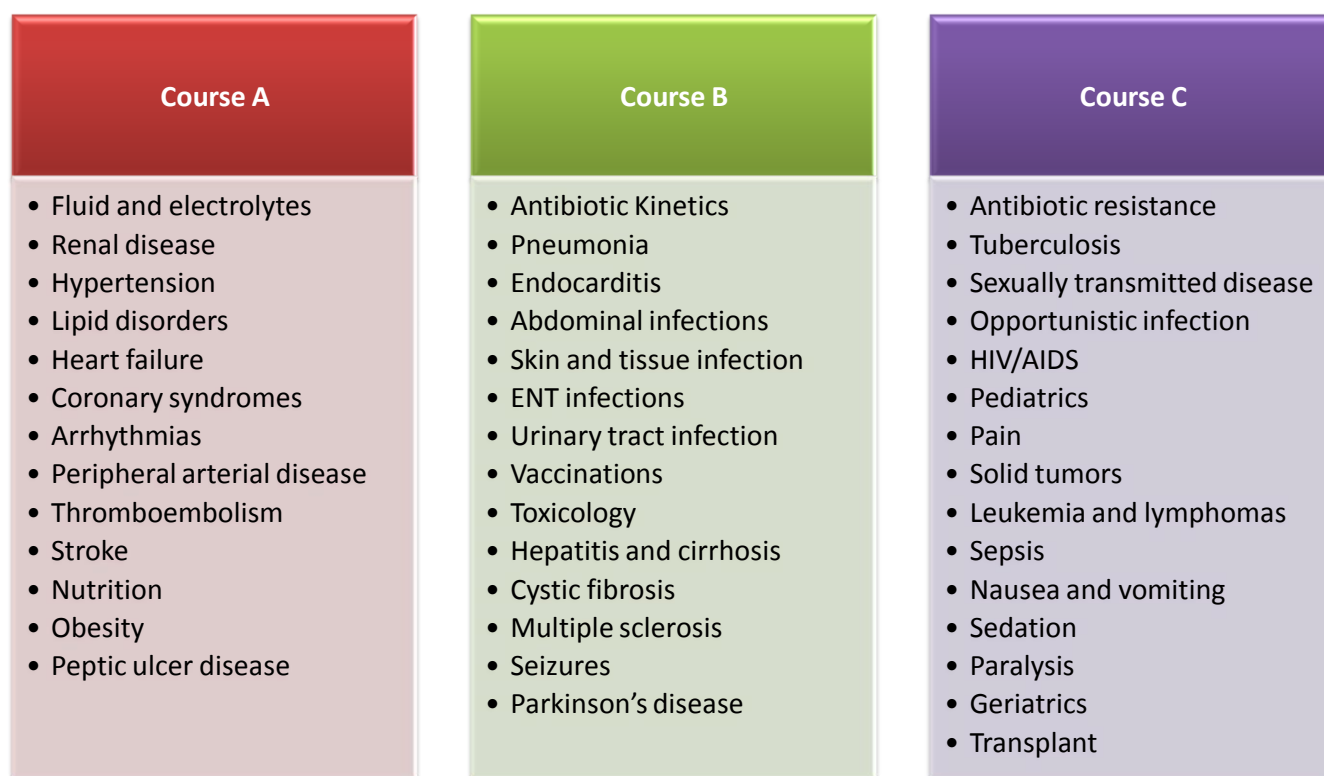


Figure 2: Major topic areas in each therapeutics course evaluated

Author Name Heart Failure - Exam II		
Recall	Content Question: 27	Answer: C
Which of the following medications is indicated in ACC/AHA stage B heart failure? a) Eplerenone b) Bumetanide c) Carvedilol d) Digoxin		
Application	Content Question: 13	Answer: A
Which of the following is a contraindication for spironolactone? a) Serum creatinine=3.0mg/dL b) Serum potassium=3.5mEq/L c) Resting heart rate=68bpm d) Blood pressure=130/85mmHg		
Analysis	Content Question: 20, 27	Answer: B
A 70 year-old female (ER weight: 80 kg) presents with to the ER with progressively worsening SOB and DOE for 24 hours. She reports no dizziness. Physical exam reveals bilateral crackles and 3+ bilateral peripheral edema. The patient was seen by her cardiologist two days ago for her regular checkup (office weight: 70 kg). Current medications include furosemide 40 mg PO daily (no recent changes), lisinopril 20 mg daily (no recent changes), digoxin 0.125 mg PO daily (no recent changes), and metoprolol XL 50 mg PO daily (increased from 25mg by cardiologist at last visit two days ago). ER vitals include: BP 140/94 mm Hg, HR 84 bpm, RR 20, and temp 98.6 F. Labs include: Na 138 mEq/L, K 4.4 mEq/L, Cl 101 mEq/L, CO ₂ 24 mEq/L, BUN 12 mg/dL, SCr 1.1 mg/dL, BS 119 mg/dL, and serum digoxin 1.2 ng/mL. Her last echocardiogram was two months ago (EF=38%). Which of the following is the <u>BEST</u> approach to manage this patient's chief complaint? a) Decrease lisinopril to 10 mg PO daily b) Increase furosemide to 80 mg PO daily c) Decrease metoprolol XL to 25 mg PO daily d) Increase digoxin to 0.25 mg PO daily		

Figure 3: MCQ submission form with sample questions.

Note: The recall question only requires knowledge that beta blockers are indicated in stage B. The application question requires interpretation of each parameter in view of a list of contraindications. The analysis question requires interpretation of multiple pieces of data, assessment of current status of problem, and evaluation of current therapy before determining the best approach to management.

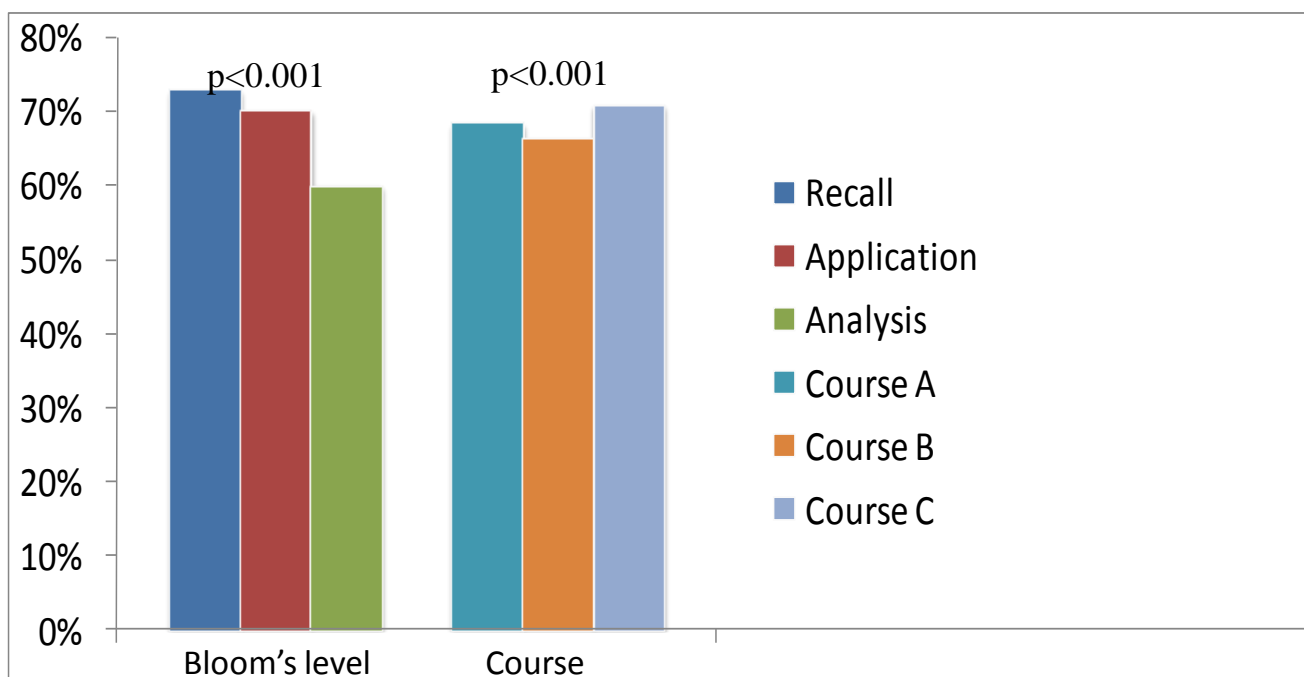


Figure 4: Overall MCQ Performance by Bloom's Level and Course

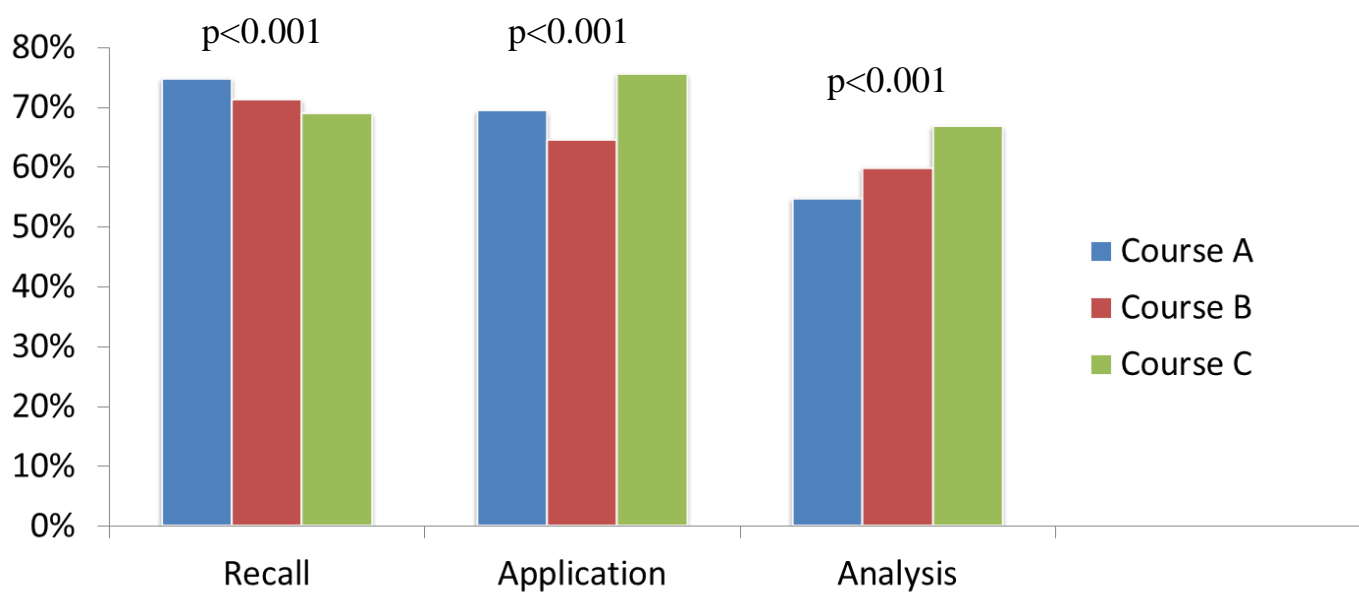


Figure 5: Student Performance (%) by MCQ Type

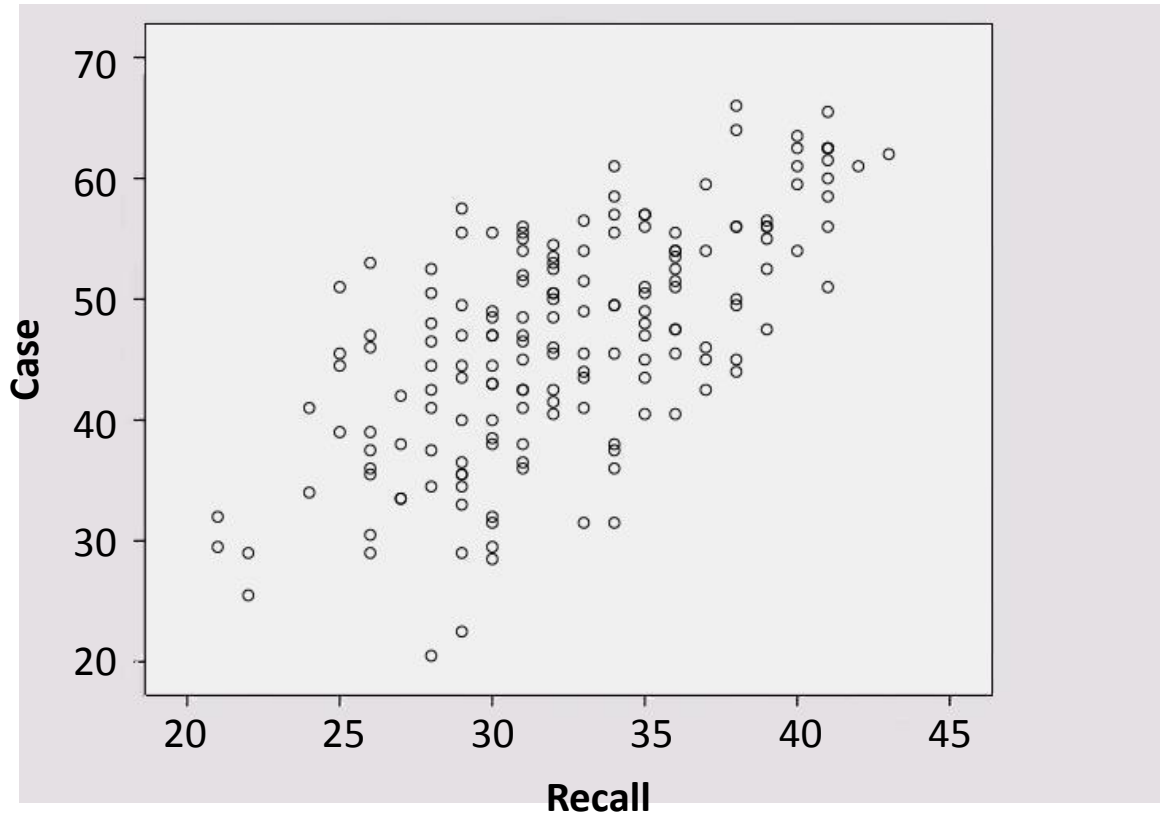


Figure 6: Scatter plot of Recall: Case Correlation