

TEACHERS' TOPIC

Test-Enhanced Learning in an Immunology and Infectious Disease Medicinal Chemistry/Pharmacology Course

Marcy Hernick, PhD, BPharm

Appalachian College of Pharmacy, Oakwood, Virginia

Submitted January 16, 2015; accepted June 22, 2015; published September 25, 2015.

Objective. To develop a series of active-learning modules that would improve pharmacy students' performance on summative assessments.

Design. A series of optional online active-learning modules containing questions with multiple formats for topics in a first-year (P1) course was created using a test-enhanced learning approach. A subset of module questions was modified and included on summative assessments.

Assessment. Student performance on module questions improved with repeated attempts and was predictive of student performance on summative assessments. Performance on examination questions was higher for students with access to modules than for those without access to modules. Module use appeared to have the most impact on low performing students.

Conclusion. Test-enhanced learning modules with immediate feedback provide pharmacy students with a learning tool that improves student performance on summative assessments and also may improve metacognitive and test-taking skills.

Keywords: test-enhanced learning, retrieval practice, curriculum, assessment, medicinal chemistry, metacognition

INTRODUCTION

Repetition is an important component of learning. However, as pharmacy schools move increasingly toward integrated courses and blocked curricula, the number of exposures or repetitions that students get to a particular subject is decreasing. Additionally, the spacing of repetitions is an important determinant of the length of time that information will be retained—a phenomenon known as “spacing effect.”¹⁻⁶ In general, longer spacing intervals between repetitions lead to longer retention intervals, as indicated by test performance on recall and recognition questions wherein the optimal spacing interval between repetitions is equal to ~10-20% of the desired retention or testing interval.¹⁻⁷

In the Appalachian College of Pharmacy curriculum, the pharmacology and medicinal chemistry lectures for a particular subject are generally held on consecutive days (ie, spacing interval=1 day), which translates into a retention or testing interval of approximately one week. Students need to retain course information longer than one week, however, so more frequent exposures to material within courses, as well as increased spacing intervals

between those exposures, are needed to tackle the challenges presented by integrated and blocked curricula.

Active learning is one technique that enhances student performance and learning outcomes.⁸⁻¹¹ Types of active learning that are used in pharmacy education include cooperative learning, problem-based learning, team-based learning, case-based learning, and ability-based education and assessment-as-learning.¹² Although testing traditionally has been used in education as a means to assess student learning, research has shown that testing also can be used to promote learning and long-term knowledge retention.¹³⁻¹⁶ The ability of testing to promote learning and knowledge retention is known as “testing effect” or “retrieval practice.”¹⁴⁻¹⁹ The use of test-enhanced learning has been an active area of psychological and educational research over the past several years. Results from these studies demonstrate that, compared to repeated studying, retrieving information during testing enhances the learning process (short-term) and slows the forgetting process, leading to superior knowledge retention (long-term).^{2,16,20-22} Furthermore, providing students with immediate detailed feedback to responses enhances the testing effect compared with students who do not receive an explanation for the responses.^{2,23,24} This feedback also serves as a way to reinforce important course concepts on a larger scale than is obtained with traditional in-class active-learning

Corresponding Author: Marcy Hernick, BPharm, PhD, Appalachian College of Pharmacy, 1060 Dragon Road, Oakwood, VA 24631. Tel: 276-498-5244. E-mail: mhernick@acp.edu

assignments; thus, it should be able to artificially lengthen the spacing interval between exposures to material, thereby lengthening the retention interval.

Test-enhanced learning has been used to promote learning and long-term knowledge retention in medical,^{15,25,26} dental,^{27,28} and chemistry/science courses.^{23,29-31} Therefore, pharmacy education could be a logical extension for this approach, given the overlapping similarities of material and the need for long-term knowledge retention in pharmacy education. However, there is little work in this area.

Self-testing was used (ie, test-enhanced learning) in pathophysiology courses in pharmacy curricula.^{32,33} In these studies, students had access to optional online quizzes as a resource to prepare for course examinations with an unlimited number of attempts. Students received feedback regarding the percentage of questions answered correctly and were able to view questions with incorrect responses (without identification of the correct response). Use of the practice quizzes led to a significant improvement in student performance on course examinations.^{32,33} Additionally, student performance on the practice quizzes was predictive of examination performance. The results from these initial studies are promising for the use of test-enhanced learning in pharmacy education.

One additional benefit associated with retrieval practice is the promotion of metacognitive skills development.^{30,32,34-38} Metacognitive skills include self-awareness of knowledge and lack thereof, as well as the ability to develop learning strategies, monitor one's learning, and make changes to the learning processes as needed to improve learning.^{30,32,34-38} In pharmacy, developing metacognitive skills is important because they are necessary for self-directed lifelong learning.^{32,37,38}

In this study, we sought to demonstrate that test-enhanced learning could be used as a curricular tool for pharmacy education, specifically for immunology and medicinal chemistry topics, to enhance learning performance and promote development of test-taking and metacognitive skills among students.

DESIGN

The Appalachian College of Pharmacy offers a 3-year doctor of pharmacy (PharmD) program that uses a modified block schedule. Because of the rapid curriculum pace, learning tools are needed that foster long-term knowledge retention. To examine whether test-enhanced learning could be used as a means to improve student performance on summative assessments and promote knowledge retention in pharmacy education, a series of optional online learning modules was created for select topics in a team-taught integrated immunology and

infectious disease pharmacology/medicinal chemistry course during the fall semester of the first (P1) year. The class met 3 hours per day, 5 days per week, for 2 months.

Students enrolled in the course offered during year 1 (n=73) did not have access to modules and were used as the control group. Students enrolled in the course offered during the fall semester of year 2, n=73) had access to the optional modules (Table 1) through the course site in Moodle (Moodle Pty Ltd, Perth, Australia). The modules typically were released to students on the day the material was presented in lecture. In general, lectures for the medicinal chemistry topics followed corresponding pharmacology lectures. This project received approval from the Appalachian College of Pharmacy Institutional Review Board.

Modules (Table 1) were developed for 6 of the 27 topics covered during the course. Online modules were created in Moodle using the quiz feature and consisted of 18-32 multiple-choice questions (ie, A-type, K-type, "select all that apply" (SATA), and matching formats. Moodle has a built-in feature that allows instructors to provide immediate feedback to students through an overall feedback field and feedback fields for individual responses (Table 2).

The overall feedback field was used to provide explanations of the overall concepts being assessed in the questions and a rationale for the correct response(s), while individual response fields were used to provide feedback about why a specific answer was incorrect. For example, if the question required students to correctly identify the drug with a given mechanism of action, the correct mechanism of action for each distractor would be provided in the response feedback fields. As noted above, this immediate feedback to responses is a key component of test-enhanced learning to improve students' conceptual understanding of the material and promote long-term knowledge retention rather than simple memorization. The use of multiple question types and detailed feedback regarding correct or incorrect responses distinguished the modules in this study from previous studies using self-testing in pharmacy education.^{32,33}

Students were allowed an unlimited number of attempts on the modules, which were not used in the calculation of course grades. Module questions for immunology topics typically required students to answer questions on specific vaccines (eg, schedules, contraindications) and hypersensitivity reactions (eg, different types, specific diseases), as well as on general knowledge regarding immune deficiencies and autoimmunity. For medicinal chemistry topics, questions tested knowledge of drug mechanisms of action, adverse drug reactions, structural drug class recognition, prodrug activation,

Table 1. Student Performance on Modules by Topic

Topic ^a	Attempt	No. Attempts ^b	Mean (SD) ^c	Median	CIC ^d
Hypersensitivity Reactions (24)	First	63 (86%)	50.8 (19.3)	55.6	83.7
	All	172	69.4 (30.3)	75.0	95.7
	Last	63	84.8 (20.8)	95.8	93.2
	Highest	63	88.1 (18.0)	96.9	92.2
Principles of Vaccination (20)	First	60 (82%)	50.8 (23.3)	51.7	86.3
	All	143	71.5 (30.4)	83.8	94.9
	Last	60	87.4 (20.1)	95.3	92.5
	Highest	60	88.4 (20.3)	97.5	93.5
Immunodeficiency & Autoimmunity (18)	First	60 (82%)	51.6 (20.8)	56.5	82.0
	All	145	73.2 (27.5)	77.8	92.7
	Last	60	83.6 (24.2)	95.4	93.8
	Highest	60	84.8 (24.1)	96.8	94.2
Antituberculosis Agents (32)	First	66 (90%)	57.7 (28.9)	62.9	94.8
	All	193	72.9 (28.0)	84.4	95.8
	Last	66	83.3 (23.8)	92.5	95.9
	Highest	66	88.1 (18.1)	93.8	93.8
Antifungals (20)	First	66 (90%)	51.6 (27.3)	53.4	90.8
	All	222	76.4 (26.8)	87.4	93.7
	Last	66	83.9 (20.9)	92.5	91.0
	Highest	66	88.2 (20.6)	97.5	94.0
Antiparasitics (22)	First	65 (89%)	50.3 (27.2)	54.7	91.3
	All	178	72.0 (28.8)	82.4	94.6
	Last	65	82.6 (24.4)	90.9	94.6
	Highest	65	85.1 (21.5)	92.5	93.3

^aTopic of module with the number of questions listed in parentheses

^bNumber of completed attempts with the percentage of students completing attempts shown in parentheses

^cStandard Deviation (SD)

^dCoefficient of internal consistency (CIC)

metabolism, contraindications/black box warnings, and drug interactions. Questions on summative assessments during year 1 were used as the starting point for module development in year 2. Additional questions were added to the modules to increase the scope and size of material to be reinforced because the goal was to use the modules as learning tools to enhance long-term knowledge retention of important material.

Results from the modules [(eg, attempts, responses, and statistics including mean, median, standard deviation (SD), coefficient of internal consistency (CIC), standard error (SE), facility index, discrimination index (DI)] were analyzed using the built-in statistics and response features in Moodle, and exported for further analysis using Excel and KaleidaGraph (Synergy Software, Reading, PA). Moodle's statistics feature provides users with information based on 4 different categories of attempts – first, all, last, and highest. We report the mean, median, SD, CIC, and SE for each module for all categories of attempts. Additionally, results for individual question performance (ie, facility index or percentage of students with correct response, DI) were exported for analysis in KaleidaGraph using the statistics and analysis of variance (ANOVA)

functions, the latter of which were carried out using an alpha=0.05.

Questions related to the topics in this study were included as part of 3 course examinations and one quiz in each of the 2 years (note: the order of antitubercular agents and antifungals was switched in years 1 and 2.). Course examinations and quizzes were administered through ExamSoft (ExamSoft Worldwide, Inc, Boca Raton, FL) and contained a combination of multiple-choice questions (MCQ), matching, and short-answer questions, with both question order and response fields randomized.

While short-answer questions are an important component of assessing students' comprehension of course material, only MCQ and matching questions are included in this analysis. There were also MCQ/matching questions on examinations not included in the modules (1-2 per year). Because of practical limitations on the number of questions that could be included on summative assessments, only a subset of module questions was included on assessments. The types of modifications used are summarized in Table 2, and include alterations of the question stem, question type, correct response, or distractors

Table 2. Sample Module Question with Detailed Feedback and Modifications used on Assessment Questions

Module Question	Response	Feedback
Which of the following drugs used to treat tuberculosis target mycolic acid biosynthesis for their therapeutic activity? (Select all that apply)	PZA	Incorrect. Plasma Membrane & Energy metabolism: target unknown (<i>Drugs: PZA</i>) (PZA prodrug activated by pyrazinamidase)
	INH	←
	ETH	←
	EMB	Incorrect. Targets Arabinoglycan transferase.
	RIF	Incorrect. Targets RNA synthesis
General Feedback		
Drugs that target the cell wall include inhibitors of mycolic acid biosynthesis: Target: InhA (enoyl-acyl-carrier protein reductase or FAS type II) <i>Drugs: INH, ETH</i> INH is a prodrug activated by KatG; ETH is a prodrug activated by EthA; both KatG and EthA are catalase peroxidase enzymes		
Type of Modification	Examples	
Question type	Conversion between multiple A-type MCQ and matching questions, Conversion between select-all-that-apply and K-type MCQ	
Correct Response(s)	Switching between drug names and drug structures, changing identity of correct response (eg, selecting another drug in the same drug class), changing the number of correct responses	
Distractor(s)	Switching between drug names and drug structures, changing identity of distractor(s), changing the number of distractors	
Stem	Changing the options for K-type questions, changing identity of drug or enzyme (with similar drug interactions), asking mechanism of action question at different level (eg, inhibit cell wall vs peptidoglycan synthesis), flipping stem/response	

Pyrazinamide=PZA, isoniazid=INH, ethionamide=ETH, ethambutol=EMB, rifampin=RIF

(number, identity, type). Results from summative assessments were analyzed using the item analysis feature in ExamSoft, and were exported for additional analysis in Excel and KaleidaGraph. Since Moodle scores all SATA and matching questions with partial credit, all values reported in Table 4 and Appendix 1 reflect question response rates with partial credit for comparison purposes. For each examination, the difference between student performance on module-related questions and all examination questions was calculated [difference=% score (module-related) questions – % score all questions]. Differences across all 3 examinations were averaged for each student. Negative percentages reflect scores on module-related questions that were lower compared to all examination questions, while positive scores indicate scores on module-related questions higher than all examination questions.

Because identical assessments were not used in years 1 and 2, we estimated the level of difficulty for each assessment by calculating the mean of the facility indices for module questions that closely “matched” assessment questions based on first attempts to ensure assessments in both years were of comparable difficulty. To examine whether student performance on the modules was predictive of performance on summative assessments, we calculated the mean of the facility indices for module

questions that closely “matched” assessment questions based on all attempts.

EVALUATION AND ASSESSMENT

Students in years 1 and 2 had comparable GPAs and PCAT scores at the time of admission (Table 3). Comparison of overall class performance on all summative assessments (all quiz and examination questions in the course from all course faculty members) in year 1 (mean 75.3%) and year 2 (mean 74.4%, $p=0.58$) revealed a modest and insignificant decrease in performance for students in year 2. Similarly, class performance on a subset of roughly 134 matched (nonshort answer) questions from all course faculty members used on summative assessments in years 1 and 2 was comparable (mean difference=-0.7%, median difference=0%). Thus, there were no significant differences between the students in the 2 cohorts, which supports the use of students in year 1 as a suitable control group.

The majority of students enrolled in the course during year 2 accessed and attempted the optional modules (82%-90%, Table 1) for a total of 166-251 attempts per module (~3-4 per user), which was largely comprised of complete attempts 143-222 per module (~2.4 to 3.4 per user). There were 8-24 partial attempts or invalid attempts per module, where students submitted responses before

AU1 Table 3. Admissions Statistics of Students from Control Cohort (Year 1) and Test Cohort (Year 2)

Statistic	GPA ^a		SGPA ^a		MGPA ^a		PCAT ^b	
	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
N	73	73	73	72	73	72	73	73
Mean	2.90	2.99	2.69	2.80	2.89	2.86	51.3	50.5
Median	2.76	2.93	2.54	2.66	2.90	2.94	50	54
SD	0.47	0.39	0.55	0.53	0.70	0.65	15.9	17.5
<i>p</i> value ^c	0.19		0.20		0.83		0.80	

^aGrade point average (GPA), science grade point average (SGPA), math grade point average (MGPA) are all out of a 4.0 scale

^bPharmacy College Admission Test (PCAT) percentile ranking (range 1-99)

^c*p* values between years 1 and 2 for GPA, SGPA, MGPA, and PCAT scores were calculated using the analysis of variance (ANOVA) function in KaleidaGraph with an alpha=0.05

answering all questions and gained access to the answer key. In general, modules for medicinal chemistry topics offered later in the course were accessed more than the modules for immunology topics offered earlier.

A summary of student performance on modules is shown in Table 1. Student performance on first attempts (mean: 50.3%-57.8%) was much lower than the performance on last attempts (mean: 82.6%-87.4%) or all attempts (mean: 69.4%-76.4%). This result was expected since most students used the modules as tools for studying rather than as a means of self-evaluation or self-testing after studying. Overall scores on the modules improved with repeated attempts resulting in increases in both the mean (82.6%-87.4%) and median (90.9%-95.8%) scores for last attempts. Coefficient of internal consistency (CIC) values for most modules was >90% regardless of the attempt category (first, last, highest, all).

Student performance on module questions improved with repeated attempts (Table 1), and this improvement in performance was observed across all topics and question types examined (Appendix 1, Tables A1 and A2). Student performance based on question type is provided in Appendix 1 (Table A1). All question types had good discrimination ability, as indicated by the DI (>50% for all question types), and student performance on all question types improved by ~25%-35% with repeated attempts.

We also examined whether the improvement in performance on module questions translated to improved performance on summative assessments. The summary in Table 4 includes results for total questions used in each year, as well as results for a subset of “matched” questions that were included on assessments in both years. The estimated difficulty and predicted performance for assessments (total and matched questions) in years 1 and 2 were comparable, indicating the assessments used in both years were of similar difficulty. Additionally, the mean testing interval across all module-related topics on summative assessments was comparable for both years (year 1: 5.8 days;

year 2: 5.3 days). Student performance on module-related questions across all assessments in year 1 closely matched the performance predicted from the modules in year 2 (predicted mean: 73%, actual mean: 75.8%). This finding further supports the use of students in year 1 as the control group. In contrast, student performance in year 2 exceeded that predicted from the module questions (predicted mean: 72.4%, actual mean: 81.9%), and was higher than student performance in year 1 (*p*=0.056). The observed scores in year 2 were more similar to the results on last/highest module attempts and were consistent with improved learning.

Results for quizzes and examinations were also analyzed separately. We found a significant improvement in student performance on examination questions (Table 4) among students who had access to modules compared with the control group without module access. Comparison of student performance on examinations revealed a significant enhancement of learning in year 2 (year 1 mean: 74.4%; year 2 mean: 83.2%, *p*=0.011). The performance gap is increased further if the results for anti-parasitics (1 day testing interval) are excluded from the analysis (year 1 mean: 72.5%; year 2 mean: 84.7%, *p*=0.0070). Only one module-related topic was included on both a quiz and examination in both years. Students in year 1 scored well on the quiz (mean 81.6%, testing interval 7 days) and then proceeded to perform poorly on the same (quizzed) material on the examination (mean 65.8%, testing interval 13 days). In contrast, year 2 students performed close to the predicted mean on the quiz (mean 71.2%, testing interval 6 days) and performed better than predicted on the same (quizzed) material on the examination (mean 88.6%, testing interval 9 days). A breakdown of student performance on examinations by topic and question type is shown in Appendix 1 (Tables A1 and A2).

In attempts to improve long-term retention of course material, a portion of the final examination in year 2 was cumulative. The majority of questions related to “old”

Table 4. Student Performance on Summative Assessments

Assessment	Parameter ^a	Year 1 ^b (Total, Matched %s)	Year 2 ^b (Total, Matched %s)	p value ^{b,c} (Total, Matched)
All assessments	No.	(27, 17)	(44, 16)	0.056, 0.072
	Estimated difficulty	54.8, 53.9	53.2, 53.4	
	Predicted performance	73.0, 72.5	72.4, 72.5	
	Mean	75.8, 73.1	81.9, 82.3	
	Median	79.5, 79.5	83.2, 86.0	
	Testing interval	5.8 days	5.3 days	
Examinations only	No.	(22, 15)	(39, 14)	0.011, 0.050
	Estimated difficulty	53.8, 55.4	53.5, 55.4	
	Predicted performance	73.0, 73.1	72.0, 73.1	
	Mean	74.4, 74.7	83.2, 84.9	
	Median	77.4, 81.0	86.5, 89.2	
	Testing interval	5.6 days	5.2 days	
Quiz only	No.	(5, 0)	(5, 0)	N/A
	Estimated difficulty	59.8	51.1	
	Predicted performance	73.0	75.8	
	Mean	81.6	71.2	
	Median	81.1	66.0	
	Testing Interval	7 days	6 days	
Examinations without antiparasitics	No.	(14, 8)	(27, 7)	0.0070, 0.14
	Estimated difficulty	54.4, 57.9	53.6, 57.9	
	Predicted performance	72.1, 72.2	71.3, 72.2	
	Mean	72.5, 73.5	84.7, 86.4	
	Median	74.0, 80.0	89.2, 93.0	
	Testing interval	6.8 days	9 days	

^aNo.=number of nonshort answer questions (Total, Matched) related to module topics on the assessment(s). Total refers to the total number of questions related to module topics on the assessment(s), while Matched refers to similar/identical questions on summative assessments in year 1 and year 2. Predicted difficulty reflects the mean facility index of the corresponding module questions for first attempts. Predicted performance reflects the mean facility index for the corresponding module questions for all attempts. Testing interval is the time between the lecture and assessment. Mean and median were calculated using the response rates for individual questions on summative assessments

^bTotal refers to the total number of questions related to module topics on the assessment(s), while Matched refers to similar/identical questions on summative assessments in year 1 and year 2

^cp values for the question response rates (Total, Matched) between years 1 and 2 were calculated using the analysis of variance (ANOVA) function in KaleidaGraph with an alpha=0.05

material were short answer questions. However, 2 MCQ questions related to hypersensitivity and vaccination principles were included on the examination (testing interval=42 days). The correct response rates to these questions were 77% and 84%, which is comparable to the response rates for similar questions on examination 1 (testing interval=5.5 days). These results are consistent with retention of the information because the majority of students (66/73) did not reattempt the modules prior to the final examination.

Because the improved performance on examinations could be attributed to an overall enhanced performance of the students in year 2 across all course topics (module and nonmodule related), we compared student performance on examination questions for module related topics with their performance on nonmodule related topics. For this analysis, we calculated the difference in performance

between module-related questions and all (nonshort answer) questions on each examination for every student, and then calculated the average difference across all examinations for each student (Figure 1 and Table 5). A positive value indicates the student performed better on the questions from module-related topics compared with overall assessments, and a negative value indicates that the student performed worse on questions from module-related topics compared to overall assessments. We define positive as a >5% increase over the total assessment and negative as a >5% decrease over the total assessment. A difference of <±5% was considered to be modest, or neutral. The overall class performance in year 1 was negative, with a mean difference of -5.6%, while the class performance in year 2 was positive with a mean difference of +2.3% for a net change of +7.9%. This significant positive shift in the average examination difference

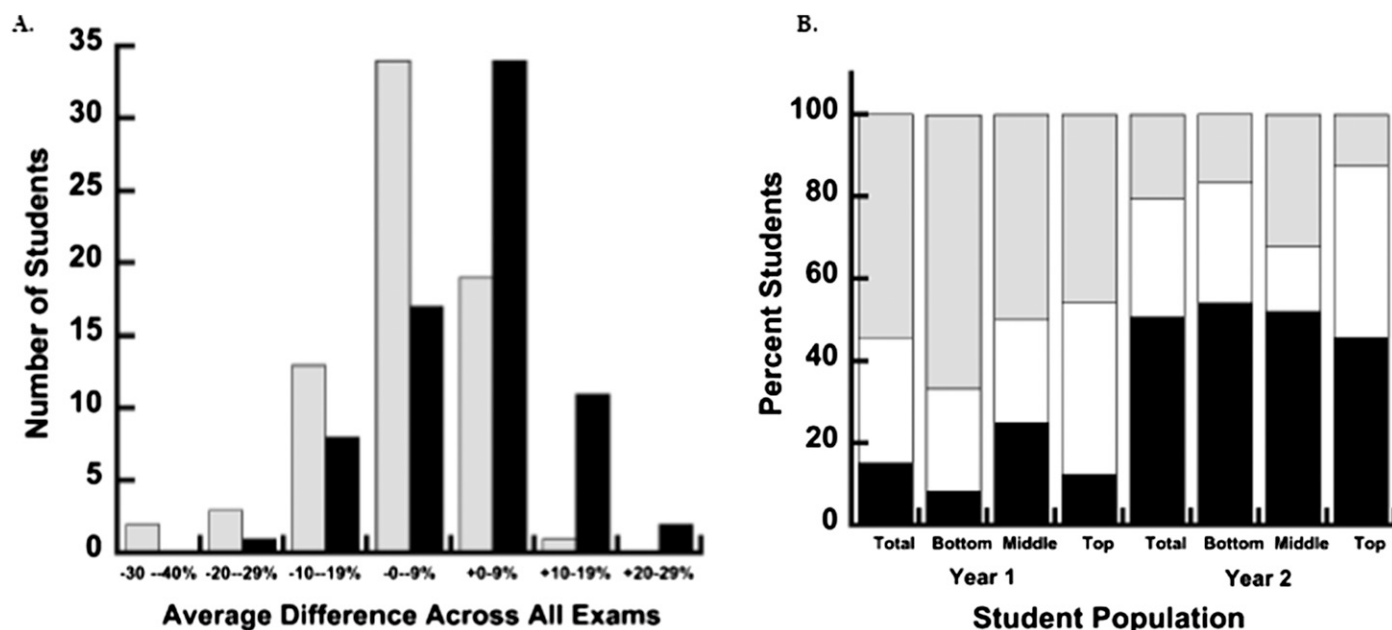


Figure 1. Comparison of the average difference between student performance on examination questions related to module topics vs all examination questions. (A) Results for students in year 1 (no modules) are represented by gray bars and students in year 2 (with access to modules) are represented by black bars. (B) Population breakdown by final course grade. (Total, top 33%, middle 33%, bottom 33%). The percent of students in each group with a positive (black), neutral (white), negative (gray) performance are shown for year 1 and year 2.

profile for students year 2 compared to year 1 is shown in Figure 1A. The results in Table 5 reveal that the majority of students in year 1 (39/72=54.2%) had a negative performance on module-related examination questions with

a mean difference of -12.2%, including 18/72 (25%) students who had high negative performances (> 10% change). In contrast, significantly fewer students had positive performances (11/72=15.3%) with a mean

Table 5. Summary of Average Examination Difference Based on Student Population

Performance ^a	Population ^b	No. Students ^c			Mean Difference (SD) ^d		
		Year 1	Year 2	Change ^e	Year 1	Year 2	Change ^e
Overall	Total	72	73	+1	-5.6 (9.3)	2.3 (8.9)	+7.9
Positive	Total	11 (1)	37 (13)	+26 (+12)	6.8 (1.7)	9.3 (4.8)	+2.5
	Bottom	2 (0)	13 (5)	+11 (+5)	7.0 (0.8)	10.4 (6.5)	+3.4
	Middle	6 (1)	13 (6)	+7 (+5)	6.7 (2.1)	10.2 (4.1)	+3.5
	Top	3 (0)	11 (2)	+8 (+2)	6.8 (1.8)	7.0 (2.0)	+0.2
Neutral	Total	22	21	-1	-0.04 (1.98)	-0.7 (2.6)	-0.74
	Bottom	6	7	+1	0.5 (1.7)	-0.6 (3.2)	-1.1
	Middle	6	4	-2	-0.07 (2.0)	-1.75 (2.0)	-1.7
	Top	10	10	0	-3.6 (2.2)	-0.3 (2.4)	+3.3
Negative	Total	39 (18)	15 (9)	-24 (-9)	-12.2 (7.1)	-10.6 (3.9)	+1.6
	Bottom	16 (9)	4 (1)	-12 (-8)	-11.6 (4.6)	-8.1 (3.0)	+3.5
	Middle	12 (3)	8 (6)	-4 (+3)	-10.1 (7.3)	-11.6 (4.1)	-1.5
	Top	11 (6)	3 (2)	-8 (-4)	-15.4 (9.0)	-11.6 (4.1)	+3.8

^aMean difference between student performance on module-related questions (MCQ, matching) and all questions on the assessment. "Positive" indicates performance on module-related questions $\geq 5\%$ higher than overall score on the assessment. "Neutral" reflects an average difference of $-5\% < x < +5\%$. "Negative" indicates performance on module-related questions $\geq -5\%$ lower than overall score on the assessment

^bStudents were divided into 3 different populations based on final course grades: Top 33%, Middle 33%, and Bottom 33%

^cNumber of students in each population with indicated performance. Number of students with effect $> 10\%$ change shown in parentheses

^dMean examination difference and standard deviation for students in the indicated population

^eDifference Year 2–Year 1

difference of +6.8%, and only one student had a high positive performance (>10% change).

After the introduction of modules in year 2, a large increase in the number of students with positive performances occurred, from 11 to 37 (15.3% to 50.7%), along with a large decrease in the number of students with negative performances, from 39 to 15 (54.2% to 20.5%). This included a significant increase in the number of high positive performances, from 1 to 13 (1.4% to 17.8%), and a decrease in the number of high negative performances, from 18 to 9 (25% to 12.3%). The improved performance on examination questions in year 2 was selective for module-related topics compared with nonmodule topics (Figure 1, Table 5) supporting that the improvement in performance was attributed to student use of the modules and not simply superior overall class performance in year 2.

Although there was not a direct correlation between test score and the total number of attempts on modules, we examined the results in more detail to determine if we could identify trends or factors associated with successful use of the modules for enhancing student learning. To this end, we compared the mean differences on examinations to cohorts identified based on final course grade, GPA, SGPA, MGPA, and PCAT score. Results based on final course grade (top 33%, middle 33%, bottom 33%) are shown in Figure 1B and Table 5. Figure 1B shows the percentage of students with positive performances (black shading) increased in year 2 for all cohorts, while the percentage of students with negative performances (gray shading) decreased in year 2 for all cohorts. Closer examination of these results (Table 5) revealed the population with the greatest increase in positive performances and greatest decrease in negative performances in year 2 was the bottom 33% of the class. The number of students in this cohort with a positive performance increased, from 2 to 13 (8.3% to 54.2%) in year 2, including an increase in the number of high positive performances, from 0 to 5 (0 to 20.8%). The number of students in this cohort with a negative performance decreased, from 16 to 4 (66.7% to 16.7%) in year 2, including a decrease in the number of high negative performances, from 9 to 1 (37.5% to 4.2%). This analysis of final course grades indicates the student population that benefited the most from the modules is the bottom 33%, as this population had the greatest increase in the number of positive performers and greatest decrease in the number of negative performers (Figure 1B, Table 5). This finding is consistent with previous studies that have shown that low performers benefit the most from active learning.^{8,9,29}

The mean examination difference results were also compared with student admission statistics (Table 6) to identify cutoff values that could be used to distinguish

high positive and high negative performers. Attempts to use GPA, SGPA, or PCAT as a single predictor were challenging, as the cutoff values were either too narrow and did not capture enough high positive or high negative students, or they were too broad and encompassed too much of the total student population to be of any value. For example, a GPA cutoff of >2.8 yielded a cohort that captured 12/13 high positive performers (mean difference +4.7) and excluded 6/9 high negative performers. However, this cohort was comprised of 49/73 students (~68%) in the class. Similar issues arose with attempts to use SGPA or PCAT scores, so combinations of these scores were examined. The best results were obtained using a combination of SGPA and PCAT scores. A PCAT/SGPA ratio cutoff of <16 yielded a cohort of 30 students that captured 9/13 high positive performers (mean difference of +5) and excludes 8/9 high negative performers.

DISCUSSION

The results in Table 1 suggest the topics students had the most difficulty with were hypersensitivity reactions (lowest mean for all attempts) and antiparasitics (lowest mean for first and highest attempts). However, the difficulty for antiparasitics may be overestimated as students only had roughly one day to use the modules prior to the final examination.

Similar to other studies examining the use of self-testing in a pathophysiology course,^{31,33} student performance on the modules in these studies was predictive of student performance on examinations. This is an important finding as it may help enable faculty members and students accurately gauge student preparedness for examinations and identify areas of weakness. Faculty members can use the module results to identify areas for curricular and examination improvement and the examination results to make improvements to the modules. Students can use the feedback they gain from the modules to gauge their preparedness for examinations, focus their study efforts (metacognitive skill), and improve their test-taking abilities.

Test-enhanced learning promotes long-term knowledge retention,¹³⁻²² while the length of time that the information is retained is dependent upon the spacing between repetitions.¹⁻⁶ Because spacing interval between repetitions in our study was one day, the spacing effect predicted the information would be retained for one week or 7 days, after which the forgetting process would occur. Although our results are preliminary, 2 pieces of data are consistent with predictions of knowledge retention and the spacing effect. First, response rates to comprehensive examination questions in year 2 (average testing interval=42 days) were comparable to the response rates to

Table 6. Module Effectiveness by Cohort

Item ^a	Population ^b	No. ^c	Mean (SD) ^d	No. High Positive ^e	No. High Negative ^e
GPA	Top (3.07–4.0)	25	4.7 (8.8)	7	1
	Middle (2.81–3.06)	24	4.7 (8.0)	5	2
	Bottom (2.19–2.79)	24	-2.5 (8.1)	1	6
	Cutoff: >2.8 (N=49)			12/13	3/9
SGPA	Top (2.95–4.0)	24	4.1 (7.8)	5	1
	Middle (2.52–2.89)	23	2.6 (9.4)	5	3
	Bottom (1.98–2.50)	25	0.3 (9.5)	3	5
	Cutoff: >2.5 (n=47)			10/13	4/9
PCAT	Top (61-93)	24	2.0 (6.6)	2	2
	Middle (43-59)	23	1.6 (9.8)	4	4
	Bottom (11-40)	25	2.9 (10.0)	7	3
	Cutoff: < 60 (n=48)			11/13	7/9
PCAT/GPA	Top (21.1–35.6)	26	-0.7 (8.3)	2	5
	Middle (13.9–21.0)	20	3.7 (7.9)	4	2
	Bottom (3.4–13.6)	26	3.8 (9.6)	7	2
	Cutoff: <21.1 (n=46)			11/13	4/9
PCAT/SGPA	Top (22.2–36.1)	29	0.8 (9.1)	3	5
	Middle (16.3–21.6)	14	1.9 (7.9)	1	3
	Bottom (4.1–15.9)	30	5.0 (8.7)	9	1
	Cutoff: <16 (n=30)			9/13	8/9

^aGrade point average (GPA), science grade point average (SGPA), and Pharmacy College Admission Test (PCAT). GPA and SGPA are out of a 4.0 scale

^bFor each item, students in the class (N=73 total) were grouped into one of 3 populations (top, middle, bottom). The ranges of scores for students in a given cohort are shown in parentheses

^cNumber of students in given cohort

^dMean average difference of performance on module topic questions vs entire examination

^eNumber of students in cohort with a positive mean examination difference > +10% and the number of students in cohort with a negative mean examination difference of <-10%

module-related questions on the initial examination (average testing interval=5.5 days), indicating that little to no forgetting occurred.

The second piece of data stems from topic results tested with a quiz and examination. Students in year 1 performed well on module-related questions on the quiz (mean 81.6%, testing interval=7 days), but performed poorly when retested on the same material just 6 days later (mean 65.8%), as predicted based on the spacing effect and forgetting curve. In contrast, students in year 2 achieved higher scores on the retest (mean 88.6%, testing interval 9 days) compared with the initial quiz (mean 71.2%) 3 days earlier. These results suggest the modules may offer a means to introduce “repetitions” outside of the classroom via an active process that artificially extends the spacing interval, thereby lengthening the retrieval interval. The small number of questions used to gauge knowledge retention is a major limitation, however, and, therefore, more robust studies are needed.

Results from short testing intervals (eg, quizzes, examination questions on antiparasitics) suggest that any benefit gained from using modules is lost if the testing

interval between the lecture and assessment is too short (eg, 1-2 days), as these circumstances tend to favor memorization of materials rather than conceptual understanding. Consequently, the modules are most effective at enhancing learning when they are used in a time frame that allows for repeated attempts over long time intervals (ie, multiple days–weeks). This finding supports the use of “reading days” in the curriculum to allow students sufficient time to digest and understand material presented.

The improvement in student performance on examinations could be attributed to an enhanced conceptual understanding of the material and/or improved test-taking and metacognitive skills. Metacognitive skills are associated with the academic success of high performing students, while low performing students tend to lack these skills.^{35,36,38}

Interestingly, cohort analysis of module effectiveness based on admissions statistics (Table 6) identified students with low PCAT/science grade point average (SGPA) ratios as the group helped most by the modules. This finding is interesting because it may support that improved test taking and metacognitive skills contribute

to the observed improvement in examination performance. If one views SGPA as an indicator of a student's intellectual capability and PCAT scores as a measure of test-taking skills, then a low PCAT/SGPA ratio could reflect that the student is intellectually capable of understanding the course material, but can underperform based on poor test-taking or metacognitive skills. In this scenario, the extra practice answering questions gained by using modules improves the student's test-taking skills, resulting in superior examination performance. This finding could have important implications for standardized testing (ie, NAPLEX). Finally, the finding that modules may promote the development of metacognitive skills in students could help explain why students who benefit the most from modules are the low-performing students (based on course grade and PCAT/SGPA), as high-performing students likely already have these skills.

SUMMARY

The results from these studies support the feasibility of using test-enhanced learning in pharmacy education to improve student performance on examinations. Importantly, students in the bottom 33% of the class (based on final course grade) and those with low PCAT/SGPA ratios most benefit from the use of test-enhanced learning. Additionally, test-enhanced learning may improve student test-taking skills, which could have important implications for standardized testing (ie, NAPLEX), and promote the development of metacognitive skills necessary for academic success and self-directed, life-long learning.

ACKNOWLEDGMENTS

The author would like to thank Veronica Keene for providing the student admissions statistics data used in this study, as well as students Savannah Horn, Vincent Nguyen, and Jonathan Ross for their assistance with preparation of the tables.

REFERENCES

1. Cepeda NJ, Vul E, Rohrer D, Wixted JT, Pashler H. Spacing effects in learning: a temporal ridge of optimal retention. *Psychol Sci*. 2008;19(11):1095-1102.
2. Pashler H, Rohrer D, Cepeda N, Carpenter S. Enhancing learning and retarding forgetting: choices and consequences. *Psychon Bull Rev*. 2007;14(2):187-193.
3. Rohrer D, Pashler H. Increasing retention without increasing study time. *Curr Dir Psychol Sci*. 2007;16(4):183-6.
4. Carpenter SK. Spacing and interleaving of study and practice. In: Benassi VA, Overson CE, Hakala CM, eds. *Applying Science of Learning in Education: Infusing Psychological Science into the Curriculum*: American Psychological Association; 2014:131-141.
5. Roediger III HL, Pyc MA. Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *J App Res Mem Cog*. 2012;1(4):242-8.
6. Kornell N, Bjork RA. Learning concepts and categories: Is spacing the "enemy of induction"? *Psychol Sci*. 2008;19(6):585-592.
7. Mozer MC, Pashler H, Cepeda NJ, Lindsey R, Vul E. Predicting the optimal spacing of study: a multiscale context model of memory. *Adv Neural Info Processing Syst*. 2009;22:1321-1329.
8. Freeman S, O'Connor E, Parks JW, et al. Prescribed active learning increases performance in introductory Biology. *CBE Life Sci Educ*. 2007;6(2):132-139.
9. Walker JD, Cotner SH, Baepler PM, Decker MD. A delicate balance: integrating active learning into a large lecture course. *CBE Life Sci Educ*. 2008;7(4):361-367.
10. Armbruster P, Patel M, Johnson E, Weiss M. Active learning and student-centered pedagogy improve student attitudes and performance in introductory biology. *CBE Life Sci Educ*. 2009; 8(3):203-213.
11. Freeman S, Eddy SL, McDonough M, et al. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*. 2014;111(23):8410-5.
12. Gleason BL, Peeters MJ, Resman-Targoff BH, et al. An active-learning strategies primer for achieving ability-based educational outcomes. *Am J Pharm Educ*. 2011;75(9):Article 186.
13. Roediger HL, Karpicke JD. Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science*. 2006;17(3):249-255.
14. McDaniel MA, Roediger HL 3rd, McDermott KB. Generalizing test-enhanced learning from the laboratory to the classroom. *Psychon Bull Rev*. 2007;14(2):200-206.
15. Larsen DP, Butler AC, Roediger HL, 3rd. Test-enhanced learning in medical education. *Med Educ*. 2008;42(10):959-966.
16. Butler AC. Repeated testing produces superior transfer of learning relative to repeated studying. *J Exp Psychol Learn Mem Cogn*. 2010;36(5):1118-1133.
17. Roediger HL, Karpicke JD. Test-enhanced learning: Taking memory tests improves long-term retention. *Psychol Sci*. 2006; 17(3):249-255.
18. Roediger HL, Agarwal PK, McDaniel MA, McDermott KB. Test-enhanced learning in the classroom: long-term improvements from quizzing. *J Exp Psychol Applied*. 2011;17(4):382-395.
19. McDaniel MA, Anderson JL, Derbish MH, Morrisette N. Testing the testing effect in the classroom. *Euro J Cog Psychol*. 2007; 19(4-5):494-513.
20. Rohrer D, Pashler H. Recent research on human learning challenges conventional instructional strategies. *Educational Researcher*. 2010;39(5):406-412.
21. Karpicke JD, Roediger HL, 3rd. Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *J Exp Psychol Learn Mem Cogn*. 2007; 33(4):704-719.
22. Agarwal P, Bain P, Chamberlain R. The value of applied research: retrieval practice improves classroom learning and recommendations from a teacher, a principal, and a scientist. *Educ Psychol Rev*. 2012;24(3):437-448.
23. Wojcikowski K, Kirk L. Immediate detailed feedback to test-enhanced learning: an effective online educational tool. *Med Teach*. 2013;35(11):915-919.
24. Wiklund-Hornqvist C, Jonsson B, Nyberg L. Strengthening concept learning by repeated testing. *Scand J Psychol*. 2014; 55(1):10-16.
25. Cantillon P. Do not adjust your set: the benefits and challenges of test-enhanced learning. *Med Educ*. 2008;42(10):954-956.

American Journal of Pharmaceutical Education 2015; 79 (7) Article 97.

26. Larsen DP, Butler AC, Lawson AL, Roediger HL 3rd. The importance of seeing the patient: test-enhanced learning with standardized patients and written tests improves clinical application of knowledge. *Adv Health Sci Educ Theory Pract.* 2013;18(3):409-425.
27. Jackson TH, Hannum WH, Koroluk L, Proffit WR. Effectiveness of web-based teaching modules: test-enhanced learning in dental education. *J Dent Educ.* 2011;75(6):775-781.
28. Baghdady M, Carnahan H, Lam EW, Woods NN. Test-enhanced learning and its effect on comprehension and diagnostic accuracy. *Med Educ.* 2014;48(2):181-188.
29. Pyburn DT, Pazicni S, Benassi VA, Tappin EM. The testing effect: an intervention on behalf of low-skilled comprehenders in general chemistry. *J Chem Educ.* 2014;91(12):2045-2057.
30. Brame CJ, Biel R. Test-enhanced learning: the potential for testing to promote greater learning in undergraduate science courses. *CBE Life Sci Educ.* 2015;14(2):14es4.
31. Dobson J, Linderholm T. Self-testing promotes superior retention of anatomy and physiology information. *Adv Health Sci Educ.* 2015;20(1):149-161.
32. Stewart D, Panus P, Hagemeyer N, Thigpen J, Brooks L. Pharmacy student self-testing as a predictor of examination performance. *Am J Pharm Educ.* 2014;78(2):Article 32.
33. Panus PC, Stewart DW, Hagemeyer NE, Thigpen JC, Brooks L. A subgroup analysis of the impact of self-testing frequency on examination scores in a pathophysiology course. *Am J Pharm Educ.* 2014;78(9):Article 165.
34. Karpicke JD, Butler AC, Roediger HL 3rd. Metacognitive strategies in student learning: do students practice retrieval when they study on their own? *Memory.* 2009;17(4):471-479.
35. Stanger-Hall KF, Shockley FW, Wilson RE. Teaching students how to study: a workshop on information processing and self-testing helps students learn. *CBE Life Sci Educ.* 2011;10(2):187-198.
36. Tanner KD. Promoting student metacognition. *CBE Life Sci Educ.* 2012;11(2):113-120.
37. Hagemeyer NE, Mason HL. Student pharmacists' perceptions of testing and study strategies. *Am J Pharm Educ.* 2011;75(2):Article 35.
38. Schneider EF, Castleberry AN, Vuk J, Stowe CD. Pharmacy students' ability to think about thinking. *Am J Pharm Educ.* 2014;78(8):Article 148.