American Journal of Pharmaceutical Education **2014; 78 (2) Article 32.**

RESEARCH

Pharmacy Student Self-Testing as a Predictor of Examination Performance

David Stewart, PharmD,[a] Peter Panus, PhD,[a,b] Nicholas Hagemeier, PharmD, PhD,[a] Jim Thigpen, PharmD,[a] and Lauren Brooks[b]

[a]Bill Gatton College of Pharmacy, East Tennessee State University, Johnson City, Tennessee
[b]College of Clinical and Rehabilitative Health Sciences, East Tennessee State University, Johnson City, Tennessee

Submitted July 9, 2013; accepted September 22, 2013; published March 12, 2014.

Objectives. To determine if student self-testing improves performance during a doctor of pharmacy course.

Methods. Students were given access to online quizzes with a large pool of randomly selected questions specific to upcoming examination content. Quizzes were electronically scored immediately upon completion and students were provided corrective feedback.

Results. Examination scores following implementation of the practice quizzes were significantly higher in all but the last testing period. The upper fiftieth percentile of students scored higher on both the practice quizzes and subsequent examinations in all but the fourth testing period.

Conclusions. Providing pharmacy students with self-testing opportunities could increase their retention of course material and provide feedback to both students and educators regarding learning, as well as provide students with a measure of their metacognition.

Keywords: pharmacy education; assessment; formative assessment; metacognition; self-testing; retrieval practice

INTRODUCTION

There are multiple ways that formative assessment has been defined.[1–5] While formative assessment typically implies assessment that occurs for the benefit of programs, teachers, evaluators, or regulatory agencies, the true meaning is much broader.[1,4] For the purpose of this manuscript we will use the inimical concept of feedback, termed "correctives," to describe formative assessment provided by the instructor or program to the individual student.[5] Correctives help students identify areas of deficiency thereby providing them a personalized formative assessment.[5] Learner-directed assessment, whether mandated or encouraged, is an integral component of developing the metacognitive skills essential for students becoming life-long self-directed learners.[2,6]

Pharmacy education is focused on addressing competencies viable well into the future of our profession,[7] and the academy should evaluate any option, including self-testing, that promotes students' individual formative assessment. To date, self-testing, also known as "retrieval practice," has received minimal attention in pharmacy education literature. One study found that medical students who self-tested had stronger first-semester academic performance than those who did not.[8]

In general, students prefer to reread notes and textbooks over retrieval practice, despite evidence that shows a significant improvement in information recall favoring retrieval practice.[9,10] One cohort of pharmacy students studying for examinations preferred to study and/or review notes or book content, to improve their confidence in the material[11]; however, this fails to provide correctives to the students. These same respondents perceived the purpose of testing as being primarily to assess the amount of material learned. This perception suggests a poor understanding of the role of testing in providing valuable formative information, which facilitates the self-directed learning process.

Self-testing has demonstrated superiority to elaborative study techniques and studying with student-generated concept maps in solidifying learning.[12] Completing frequent quizzes throughout a course can enhance the encoding and retention of concepts discussed during class.[13] Self-directed life-long learners must have good metacognitive skills (understanding what they do and do not know) in order to adequately assess their level of knowledge, yet students typically have a poor understanding of this

Corresponding Author: David W. Stewart, PharmD, Associate Professor of Pharmacy Practice, East Tennessee State University, Bill Gatton College of Pharmacy, Box 70657, Johnson City, TN. Tel: 423-439-2071. Fax: 423-439-6350. E-mail: stewardw@etsu.edu

and fail to see the value of techniques such as self-testing when studying.[10,14]

The Accreditation Council for Pharmacy Education's Standards for Doctor of Pharmacy Degree Program's Standard 11 states that teaching and learning methods should "foster the development and maturation of critical thinking and problem-solving skills. . .and enable students to transition from dependent to active, self-directed, lifelong learners."[15] Given the evidence in favor of frequent self-testing and the desire to develop lifelong learners with good metacognitive skills, this study was conducted at the Bill Gatton College of Pharmacy, East Tennessee State University, to determine whether classroom-based use/implementation of the practice retrieval concept, along with correctives, affected student examination performance. We hypothesized that student performance on subsequent examinations would improve if students used the opportunity to self-test outside of class by taking multiple-choice online quizzes to assess their comprehension of material covered in the classroom.

METHODS

This study involved first-year pharmacy students enrolled in Pathophysiology, a 17-week, 4-credit hour required course that had 4 examinations (all noncumulative) equally spaced and weighted throughout the semester. Examinations were not returned to students, thereby eliminating the risk of the content being circulated among students who have not yet taken the examination. The examinations made up 80% of the final course grade.

There were also 8 required, graded quizzes equally spaced and weighted throughout the course that accounted for the remaining 20% of the final grade. The purpose of the required quizzes was for grading and assessment of student progress in the course. In 2011, a self-testing, non-graded component was added to the course to allow students to practice retrieval of course information and provide them with formative feedback on their level of metacognition.

After the content for the upcoming examination was discussed in class, the required graded quiz was conducted. Following this quiz, a subsequent practice pool of questions (ie, the self-testing quiz) was released to the students using the university's course management software, Desire-2-Learn, version 8.3 (Kitchener, ON, Canada). Between the time the graded quiz was administered and the examination for that portion of the course was administered, students had an unlimited number of opportunities to access online self-testing. Each quiz consisted of 15 questions randomly selected by the software and students were given a maximum of 30 minutes to complete it. The software immediately graded the practice quiz upon submission, and the students could see the percentage of questions answered correctly. Furthermore, students were able to view those questions they answered incorrectly along with all potential answer choices, but the correct answer choice was not indicated. The question pool closed once the examination for that portion of the course was administered.

Following approval by the East Tennessee State University Institutional Review Board for Human Research and completion of the course, all data were extracted from the Desire-2-Learn software and organized using Microsoft Excel 2007. For the practice quizzes, the following independent variables were collected for each student: number of self-testing sessions, high and low quiz scores, scores on first and last quiz attempted, range between high and low quiz scores, and number of days before the examination during which the practice question pool was available for students to use. Undergraduate grade point average (GPA) and total score for the Pharmacy College Admissions Test (PCAT) were also included in the database.

All data analyses were completed using SPSS Statistics19 (IBM Inc, Armonk, NY). The independent variables were examined using a Pearson's correlation to determine their relationship with corresponding examination scores. A two-way repeated measures analysis of variance (ANOVA) with interaction was conducted to determine if there was any differences in the practice quiz averages compared to subsequent examination score. The 2 main effects studied were time between incorporating the quiz and subsequent examination while score represented the score for the practice quiz or the corresponding examination. The 2 main effects studied were time block relative to each examination period while score represented the percentage correct value for the practice quiz or the corresponding examination. Mauchly's test for sphericity was conducted on main effect "score" and interaction between "score" and "time." Greenhouse-Geisser correction was used to correct the F statistic. Ten paired t tests were conducted with significance set at $p<0.005$ for Bonferroni adjustment on multiple-paired comparisons.

Finally, 2 group comparisons were conducted to determine whether there were differences in student examination grades based on their previous practice quiz averages. For each examination, the examination scores and practice quiz averages were rank ordered based on the practice quiz averages prior to the examination. These rank orders were then divided into upper and lower fiftieth percentiles. Comparisons of examination scores and practice quiz averages between the upper and lower fiftieth percentiles were conducted with a 2-tailed t test for independent samples, equal variance not assumed. When variation is reported, all data are reported as arithmetic mean plus or minus the standard error of the mean. All graphs

were created using Slide Write Plus for Windows, version 5.0 (Advanced Graphics Software Inc, Encinitas, CA).

RESULTS

Seventy-nine students were enrolled in the course in the spring semester of 2011. Over 96% of these students used self-testing at least once during the semester, with 7,312 total attempts recorded. Additional descriptive statistics associated with the self-testing practice quiz pool and subsequent examination scores are reported in Table 1. The total number of attempts peaked during the period prior to the second examination and subsequently declined. The partial correlation matrix demonstrated that the overall average for the practice quizzes was consistently the first or second highest correlate to examination scores (Table 2). Neither undergraduate GPA nor composite PCAT score consistently correlated with each individual examination score to a significant extent.

A two-way repeated measures ANOVA was conducted to examine the differences between practice quiz average and subsequent examination score for each testing period. Both variables demonstrated significant differences and their interaction was also significant ($p<0.001$). These results suggest there were differences between scores for each testing period and between the practice quiz averages and subsequent examination scores. The significant interaction between these 2 variables reflected that during the first 3 of the 4 testing periods, the improvement in practice quiz averages was consistently reflected by a similar magnitude of change in the examination score (Figure 1); however, during the fourth examination interval, there was not a similar increase in the examination scores based on practice quiz average. The practice quiz taken during period 3 had the lowest average score for the period and was significantly less than that for the first period. Additionally, scores on examinations 3 and 4 were also lower compared to scores on the first examination.

The results of rank ordering the cohort into upper and lower 50th percentiles is shown in Figure 2. As expected, the upper fiftieth percentile scored higher than the lower fiftieth percentile on the practice quizzes during each of the 4 testing periods as well as for each examination, with the exception of the final examination (examination 4).

DISCUSSION

In this retrospective study of self-testing in a pathophysiology course, we found that 76 of 79 students used the opportunity (over 96% of the class) for self-assessment. While no additional credit was given for self-testing, students made 7,312 attempts. While some might argue that students took the quizzes in an attempt to memorize the questions and answers, doing so would have been almost impossible because of the large number of questions from which the software randomly selected the quiz items (n=1,342) and the fact that students were not given the correct answers if they answered incorrectly. Furthermore, if students only motivation was to ascertain the format in which the instructor asked questions, the total number of attempts probably would have been much lower.

In order for correctives to be beneficial, students must be given objective information they must apply to arrive at an answer rather than simply given the correct answer.[16] The former challenges their critical-thinking skills while the latter reinforces the less-effective study behavior that most students favor, ie, rereading notes and texts. Providing both the distracters and the correct answer choice without highlighting the correct answer was intended to force students to go back and relearn material because understanding why the incorrect answers are incorrect is just as beneficial for metacognition as knowing why the correct choice is indeed correct.[16]

The value of and parameters involved for self-testing in long-term learning has been documented in multiple experimental investigations. Similar to our data, self-testing parameters that appear to improve recall and retention include number of attempts, spacing between attempts, value of failure to retrieve information, and inclusion or exclusion of successfully retrieved material.[17–19] Many of these investigations used free recall formats to demonstrate retained information and learning.[9,12,17–20] However, other laboratory and classroom-based investigations have documented that self-testing with multiple- choice questions

Table 1. Pharmacy Students' Use of Optional Self-Administered Practice Quizzes to Prepare for Examinations (N=79[a])

In Preparation for:	No. of Questions in Pool	Days Pool Open, No.	Total Attempts, No. (Range per Student)	Practice Quiz Mean (SEM)	Examination Score
Examination 1	278	16	1695 (0-100)	80.2 (1.2)	89.3 (0.80)
Examination 2	247	14	2348 (0-150)	81.6 (1.5)	90.4 (0.82)
Examination 3	386	17	1384 (0-70)	75.4 (1.3)	82.2 (0.98)
Examination 4	431	19	1615 (0-90)	81.5 (1.0)	81.9 (0.89)

[a] The number of students completing practice quizzes during the 4 examination preparation periods varied from 70 to 77.

Table 2. Partial Correlation Matrix Comparing Variables Between Self-Attempt Practice Quizzes and Corresponding Examination Score

Practice Quiz in Preparation for:	Total Attempts	High Score	Low Score	Average Score	Last Attempt	GPA	PCAT
Examination 1 (n=76)	0.473[d*]	0.547[b*]	0.148	0.591[a*]	0.513[c*]	0.385[e*]	0.143
Examination 2 (n=77)	0.431[d*]	0.532[c*]	0.188	0.670[a*]	0.606[b*]	0.263[e*]	0.292[*]
Examination 3 (n=73)	0.373[e]	0.543[b]	0.252[*]	0.550[a*]	0.387[d*]	0.422[c*]	0.247[*]
Examination 4 (n=70)	0.181	0.232[e]	0.077	0.373[b*]	0.293[d*]	0.343[c*]	0.428[a*]

Abbreviations: GPA=undergraduate grade point average; PCAT=composite score on pharmacy college aptitude test.
Rank order of strength of correlation to corresponding exam score: a=first, b=second, c=third, d=fourth, e=fifth.
* $p<0.05$

improves long-term retention and learning for periods extending out as far as the end of the academic year.[16,21,22] An evaluation of first-year medical students validated this finding in a professional health program.[23] There may also be additional benefit to using multiple-choice self-testing questions when multiple-choice questions are the primary evaluation tool used by the course. Furthermore, in professional programs such as pharmacy, familiarity with strategies for correctly answering multiple-choice questions may result in students being better prepared for taking standardized tests such as the North American Pharmacists Licensure Examination and the Multistate Pharmacy Jurisprudence Examination.

Our results also demonstrate that PCAT score and GPA did not consistently correlate highly with examination score, lending credence to the argument that self-testing is beneficial for improving students' performance. This is significant because it makes underlying intelligence a less likely influence on our data and thereby allows us to hypothesize that higher average scores on self-testing may be directly related to examination performance. The implication of these data is that students can "gauge" their mastery of material objectively, prior to examination, thereby improving their metacognition. This could provide encouragement for them to continue or change their study methodology and/or allow the instructor to preemptively intervene, so students should be informed of this when self-testing is employed within a course.

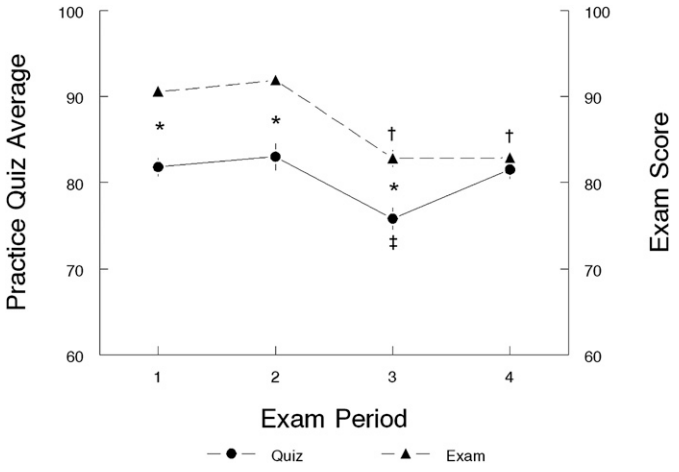

Figure 1. Class average on the self-attempted practice quizzes compared to subsequent examination score. Practice quiz question pool average (solid circles) and examination score average (solid triangles) depicted. (*) = $p<0.005$ for practice quiz average compared to corresponding exam score; (‡) = $p<0.005$ for subsequent practice quiz average compared to the first practice quiz average; (†) = $p<0.005$ for examination score compared to first examination score. (All p values significant after Bonferroni adjustment for multiple paired comparisons)

Our data were also consistent across 3 of the 4 examinations administered during this 17-week semester long course. While there are no data to provide insight into our vast differences with all correlations on examination 4, we believe that because this was a final examination and administered after the conclusion of the semester, it was completed in a different environment from that in which students completed the first 3 examinations. For example, students often "triage" examinations based on variables such as current course grade and/or anticipated final course grade, and this can affect the amount of effort they put into studying for final examinations. In support of this is the high correlation of PCAT score and GPA with examination 4, which was not observed for the first 3 examinations.

Limitations with our retrospective observations include the nonrandomized and non-controlled environment of the study. Secondary to this methodology, the number of practice questions available for each quiz increased as the semester progressed, which was simply a reflection of increased time spent writing questions for the question pool. Because there was no incentive given to students, use of the question pool was not standardized or consistent during each testing period, with a resultant spike in attempts earlier in the semester that tapered as the semester progressed.

Despite these limitations, valuable lessons were learned from this experience. Previously when practice

retrieval techniques were used in a classroom setting with student self-directed participation outside of class, the determination of such participation was by survey and self-reporting.[22] Our observations remedied these limitations by using an online course management system, Desire-2-Learn, that provided a well-designed testing platform and allowed retrospective data collection. By collating student testing data, we were able to provide multiple data points that included total number of attempts, high score, low score, average score, and last attempt, all at the individual student level.

An investigation using such an online course management system examined the benefit of self-directed, self-testing in the form of online mastery quizzes outside of the classroom on subsequent examination scores in a human anatomy class during the first year of dental education.[24] These investigators observed a significant correlation between score for the initial attempt at the online quiz and the score on the subsequent examination over the same material. We similarly found that average quiz score highly correlated with subsequent examination scores on 3 of the 4 examinations.

One question that commonly arises is whether multiple-choice questions are an appropriate assessment tool for evaluating higher levels of learning. Those types of data have not been available for retrospective comparison; however, with the advent of technology, many schools are tracking these types of data at the individual student, instructor, and even question level throughout the curriculum. With the adoption of robust electronic testing platforms with detailed assessment capabilities, investigators may be able to answer the question of whether self-testing is better suited for some levels of learning than others. Future research endeavors should focus on whether self-testing is appropriate for long-term retention of knowledge (ie, end of program outcomes), as well as on levels of learning such as by evaluating based on Bloom's taxonomy. This specific analysis addressed student performance within 1 offering of 1 course. Methodology previously employed[25] could be used to further assess whether self-testing provides benefit over other teaching techniques by comparing to cohorts who received different types of instruction.

Almost all colleges and schools of pharmacy are using active learning within their curriculum, the majority of which involves team-based or problem-based techniques.[26] Although students typically prefer active learning compared to lectures, some evaluations have shown unfavorable effects with certain active-learning techniques.[27,28] Active learning is more about engaging the student than a specific process[29]; thus, self-testing should qualify as active in the sense that it engages students and provides them correctives for their metacognitive development.

CONCLUSION

Self-testing is a largely untapped resource with strong favorable evidence for use in metacognitive skill development that is imperative for the independent lifelong learner. When we implemented student self-testing outside of class to provide pharmacy students with directed formative assessment, students used the opportunity despite the lack of any additional incentives. Similar to other investigators, we demonstrated correlation between student performance on self-testing practice quizzes and subsequent examination scores.

REFERENCES

1. Bell B, Cowie B. The characteristics of formative assessment in science education. *Sci Educ*. 2001;85(5):536-553.
2. Nicol DJ, Macfarlane-Dick D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Stud High Educ*. 2006;31(2):199-218.
3. Rolfe I, Mcpherson J. Formative assessment: how am I doing? *Lancet*. 1995;345(8953):837-839.
4. Guskey TR. Lessons of mastery learning elements mastery learning and other interventions share. *Educ Leadership*. 2010;68(2):52-57.
5. Bloom B, Hastings J, Madaus G. *Handbook on Formative and Summative Evaluation of Student Learning*. New York, NY: McGraw-Hill;1971.
6. Yorke M. Formative assessment and its relevance to retention. *High Educ Res Dev.* 2001;20(2):115-126.
7. Jungnickel PW, Kelley KW, Hammer DP, Haines ST, Marlowe KF. AACP curricular change summit supplement: addressing competencies for the future in the professional curriculum. *Am J Pharm Educ*. 2009;73(8):Article 156.
8. West C, Sadoski M. Do study strategies predict academic performance in medical school? *Med Educ*. 2011;45(7):696-703.
9. Roediger HL, Karpicke JD. Test-enhanced learning - taking memory tests improves long-term retention. *Psychol Sci.* 2006;17(3):249-255.
10. Karpicke JD, Butler AC, Roediger HL. Metacognitive strategies in student learning: do students practice retrieval when they study on their own? *Memory*. 2009;17(4):471-479.
11. Hagemeier NE, Mason HL. Student pharmacists' perceptions of testing and study strategies. *Am J Pharm Educ*. 2011;75(2):Article 35.
12. Karpicke JD, Blunt JR. Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*. 2011;331(6018):772-775.
13. Nevid JS, Mahon K. Mastery quizzing as a signaling device to cue attention to lecture material. *Teach Psychol*. 2009;36(1):29-32.
14. Karpicke JD. Metacognitive control and strategy selection: deciding to practice retrieval during learning. *J Exp Psychol*. 2009;138(4):469-486.
15. Accreditation Council for Pharmacy Education. Accreditation standards and guidelines for the professional program in pharmacy leadning to the doctor of pharmacy degree. 2011. https://www.acpe-accredit.org/pdf/FinalS2007Guidelines2.0.pdf. Accessed July 8, 2013.
16. Little JL, Bjork EL, Bjork RA, Angello G. Multiple-choice tests exonerated, at least of some charges: fostering test-induced learning and avoiding test-induced forgetting. *Psychol Sci.* 2012;23(11):1337-1344.
17. Karpickee JD, Roediger HL. The critical importance of retrieval for learning. *Science*. 2008;319(5865):966-968.

18. Bahrick HP, Hall LK. The importance of retrieval failures to long-term retention: a metacognitive explanation of the spacing effect. *J Mem Lang*. 2005;52(4):566-577.
19. Karpicke J, Roediger H. Repeated retrieval during learning is the key to long-term retention. *J Mem Lang*. 2007;57(2):151-162.
20. Karpicke JD, Roediger HL. Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *J Exp Psychol*. 2007;33(4):704-719.
21. Butler AC, Karpicke JD, Roediger HL. The effect of type and timing of feedback on learning from multiple-choice tests. *J Exp Psychol Appl.* 2007;13(4):273-281.
22. Roediger HL, Agarwal PK, McDaniel M a, McDermott KB. Test-enhanced learning in the classroom: long-term improvements from quizzing. *J Exp Psychol Appl.* 2011;17(4):382-395.
23. Larsen DP, Butler AC, Roediger HL. Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Med Educ*. 2013;47(7):674-682.
24. Lee LiM, Nagel RW, Gould DJ. The educational value of online mastery quizzes in a human anatomy course for first-year dental students. *J Dent Educ*. 2012;76(9):1195-1199.
25. Stewart DW, Panus PC, Hagemeier NE. An analysis of student performance with podcasting and active learning in a pharmacotherapy module. *Curr Pharm Teach Learn*. 2013;5(6):574-579.
26. Stewart DW, Brown SD, Clavier CW, Wyatt J. Active-learning processes used in US pharmacy education. *Am J Pharm Educ*. 2011;75(4):Article 68.
27. Persky AM. Multi-faceted approach to improve learning in pharmacokinetics. *Am J Pharm Educ*. 2008;72(2):Article 36.
28. Smythe G, Hughes D. Self-directed learning in gross human anatomy: assessment outcomes and student perceptions. *Anat Sci Ed*. 2008;1:145-153.
29. 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.