
ORIGINAL ARTICLE

Can formative quizzes predict or improve summative exam performance?*

Niu Zhang, MD, MS and Charles N.R. Henderson, DC, PhD

Objective: Despite wide use, the value of formative exams remains unclear. We evaluated the possible benefits of formative assessments in a physical examination course at our chiropractic college.

Methods: Three hypotheses were examined: (1) Receiving formative quizzes (FQs) will increase summative exam (SX) scores, (2) writing FQ questions will further increase SE scores, and (3) FQs can predict SX scores. Hypotheses were tested across three separate iterations of the class.

Results: The SX scores for the control group (Class 3) were significantly less than those of Classes 1 and 2, but writing quiz questions and taking FQs (Class 1) did not produce significantly higher SX scores than only taking FQs (Class 2). The FQ scores were significant predictors of SX scores, accounting for 52% of the SX score. Sex, age, academic degrees, and ethnicity were not significant copredictors.

Conclusion: Our results support the assertion that FQs can improve written SX performance, but students producing quiz questions didn't further increase SX scores. We concluded that nonthreatening FQs may be used to enhance student learning and suggest that they also may serve to identify students who, without additional remediation, will perform poorly on subsequent summative written exams.

Key Indexing Terms: Chiropractic; Education; Formative Assessment; Summative Evaluation

J Chiropr Educ 2015;29(1):16–21 DOI 10.7899/JCE-14-12

INTRODUCTION

Teaching faculties are interested in factors that predict academic success and facilitate learning. American College Testing (ACT) scores, grade point average, and language familiarity are widely recognized predictors of subsequent academic performance.^{1–3} It also is appreciated that student achievement improves when faculty members provide postassessment feedback.^{4–6} Therefore, it has become common for faculty to supplement traditional summative exams with formative quizzes.^{7,8} Faculty members also use quizzes to predict exam scores,⁹ and web-based self-assessments have been used to improve knowledge and test performance.¹⁰ While summative exams evaluate student knowledge or task performance at the end of instructional segments, formative assessments provide students with feedback about how they are doing along the way.⁸ Students must view formative assessments as relatively “low-stake” or “nonthreatening” for them to

be effective.^{8,11,12} Therefore, formative assessments usually are voluntary, with little or no impact on course credit.^{8,12} It is recommended that instructors should follow formative assessments with specific guidance and remedial instruction for students who demonstrate serious deficiencies in understanding, knowledge, or competence. Despite their wide use, the value of formative exams remains unclear. Haberyan,⁷ and Brothen and Wambach¹³ reported that formative assessments did not enhance overall learning outcomes, while Kibble,¹¹ Olson and McDonald,¹² and Buchanan¹⁴ reported significant improvements. Although this issue has been examined in dental¹² and medical¹⁵ curricula, we did not find similar studies in chiropractic education programs.

Therefore, we decided to evaluate the possible benefits of formative assessments in a physical examination course at our chiropractic college. This course is offered in the 3rd quarter of a 13-quarter program. We considered that student participation in developing multiple-choice quiz questions used in the formative assessments would be a valuable learning exercise. Therefore, we hypothesized: (1) The use of formative quizzes during the course would increase summative exam scores, (2) student participation in writing formative quiz questions (student generated quiz

*This paper was selected as a 2014 Association of Chiropractic Colleges – Research Agenda Conference Prize Winning Paper – Award funded by the National Board of Chiropractic Examiners

Table 1 - Study Timeline (per Term)

Week	1	2	3	4	5	6	7	8	9	10
Lectures	Head	Neck	Ear	Eye		Cranial nerves	Cerebellum	Mouth/throat	Nose/sinuses	
Labs	Head/neck	Ear	Eye	Cranial nerves		Cerebellum	Oral/nasal	Review		
FQs	FQ1	FQ2	FQ3	FQ4		FQ5	FQ6	FQ7	FQ8	
FQRs				FQR1					FQR2	
Exams					SX1					SX2

Lectures, physical findings and associated conditions; Labs, physical exam procedure instruction and practice for each topic area; FQs, formative quizzes (FQ1–8), topical area specific; FQRs, formative quiz reviews (FQR1 and FQR2) reviewing topic weeks 1–4 and 6–9; Exams, summative examinations (SX1 and SX2) at weeks 5 and 10.

questions [SGQQs]) would produce additional increases in summative exam scores, and (3) formative quiz scores would predict subsequent summative exam performance.

METHODS

Student Participants

The institutional review board of Palmer College of Chiropractic granted this educational method study an exemption from formal review. Permission was obtained from all students to use de-identified performance assessments for this study and subsequent publications.

A total of 189 3rd quarter students participated in the study across 3 separate iterations of a 3-credit physical examination Class that addressed head and neck examination procedures and related health conditions (March 2012–March 2013). Students in Class 1 created a bank of multiple choice SGQQs that subsequently were administered as 8 formative quizzes (FQs) in Class 1 and Class 2 (Table 1). The students in Class 2 were given FQs, but did not contribute to the quiz question bank. In addition, students in Class 1 and Class 2 were given formative quiz reviews (FQRs) 1 week before each of the summative exams (SXs) that were administered at the midpoint and end of the course, respectively (Table 1). Each of the FQRs reviewed the 4 FQs that preceded them and were intended to prepare the students for the SX that was to be administered the following week. While the SGQQs covered the material that would be tested in SXs, the SXs did not contain exact replicates of SGQQs. Class 3 students served as a control, taking SXs, but not producing SGQQs or taking FQs and FQRs.

Course materials, with the exception of FQs and FQRs, were equivalent for all 3 classes. In addition, the same instructor taught the 3 classes, taking care to cover the same material with equivalent class-time allocation. Demographic data (sex, age, academic degrees, and ethnicity) also were collected for the 3 classes. As shown in Table 1, FQs and FQRs were given at weeks 1–4 and 6–9, while SX1 and SX2 were administered at weeks 5 and 10, respectively.

Formative Quiz Question Bank

Class 1 students developed a bank of multiple-choice quiz questions addressing head and neck conditions, and physical exam procedures. These SGQQs were required to adhere to 2 guidelines: (1) all questions were required to have 4 multiple-choice options with only 1 correct answer

and (2) questions were to be drawn from the course-required textbook or notes. Each Class 1 student was asked to create 1 question for each of the 8 topic areas covered in the course (head, neck, ear, eye, nose, mouth, cranial nerves, and cerebellum). All SGQQs were reviewed carefully by the course instructor and validated by a knowledgeable faculty member not involved in teaching the course. Inaccurate or irrelevant questions either were modified or discarded. After evaluation and validation, all submitted SGQQs were subgrouped according to the 8 course topics to produce a question bank for the FQs.

Formative Quizzes and Summative Exams

A total of 8 FQs was administered to Class 1 and Class 2 over the 8 consecutive academic weeks of the term in which each class was enrolled (1 quiz per week, Table 1). Each quiz consisted of 6 to 26 SGQQs selected from the question bank and pertained to the topic covered in the given course week. Class 1 and Class 2 students were allowed 1 minute for each question. The FQs were administered during lab and answers were announced after students submitted their answer sheets to the instructor. Students in Class 1 and Class 2 also were given FQRs in the week before each of the two SXs (Table 1). For these reviews, students completed an online summary quiz composed of the preceding 4 FQs via Blackboard Learn (Blackboard Inc, Indianapolis, IN). This allowed students easy off-campus access. Hardcopy answer sheets were submitted in lab, after which the instructor gave open feedback concerning the correct answers. Summative exams were administered the following week to all 3 classes during lecture hours.

Data Analysis

Data were summarized and analyzed using SPSS version 22 (IBM Corporation, Armonk, NY). Statistical test assumptions were verified and *P* values less than .05 were considered significant. We applied a 1-way analysis of variance (ANOVA) with orthogonal contrasts to evaluate formative quiz effects across the 3 classes: Hypothesis 1 – Receiving formative quizzes will increase summative exam scores and Hypothesis 2 – Writing formative quiz questions will further increase summative exam scores. The ability to predict summative exam scores with formative quizzes (hypothesis 3) was evaluated with multiple linear regression using the forced entry method (copredictors of sex, age, academic degrees, and ethnicity). Age and ethnicity were collapsed to dichotomous variables

Table 2 - Demographic Data for the 3 Classes (n = 189)

Classes ^a	Class 1	Class 2	Class 3
Student count ^b	48 (100)	65 (100)	76 (100)
Sex			
Male	28 (58)	36 (55)	52 (68)
Female	20 (42)	29 (45)	24 (32)
Academic degree			
Bachelor only	42 (87)	60 (92)	74 (97)
Master or PhD	6 (13)	5 (8)	2 (3)
Age, y			
<30	41 (85)	55 (85)	69 (91)
30–40	6 (13)	8 (12)	6 (8)
>40	1 (2)	2 (3)	1 (1)
Ethnicity			
Caucasian	32 (67)	51 (78)	59 (77)
Hispanic	8 (17)	5 (8)	5 (7)
African American	4 (8)	6 (9)	5 (7)
Other	4 (8)	3 (5)	7 (9)

^a Class 1 students wrote and received formative quizzes. Class 2 students received, but did not write, formative quizzes. Class 3 (control) students neither wrote nor received formative quizzes.

^b Values in parentheses are percent of total class.

for regression analysis (age, ≤ 30 or > 30 years; ethnicity, Caucasian or minority).

RESULTS

Demographic Data and Descriptive Statistics

A total of 189 students participated in this study and demographic data are summarized in Table 2. In our sample, sexes were fairly similar in distribution between Classes 1 and 2, with a slightly greater percentage of males. This distribution was more heavily skewed toward males in Class 3. Academic degree, age, and ethnicity were markedly skewed within all classes in favor of bachelor degrees, < 30 years of age, and Caucasians. A total of 189 students participated in this study. Mean assessment scores are reported for the 3 classes in Table 3.

Between-Class Comparisons of Summative Exam Scores

Of 295 SGQQs received, 162 were discarded due to irrelevance, inaccuracy, wrong format, or redundancy. Most discards were due to question redundancy. Therefore, the formative quiz question bank consisted of 133 SGQQs.

A 1-way independent ANOVA demonstrated a moderate ($\omega^2 = .054$),¹⁶ statistically significant difference in total SX scores (SX1 + SX2) between the 3 classes (Table 4). Planned linear contrasts revealed that total SX scores for the control group (Class 3) were significantly less than those of Classes 1 and 2 – first contrast. However, the act of writing quiz questions and taking formative quizzes (Class 1) did not produce significantly higher total SX scores than only taking the formative quizzes (Class 2) – second contrast.

Formative Quizzes as Predictors of Summative Exam Scores

The capacity to predict total SX scores from total FQ scores (sum of all 8 formative quizzes) was evaluated by multiple linear regression while accounting for sex, academic degree, age, and ethnicity (Table 5). Total FQ scores were found to be statistically significant predictors of total SX scores, accounting for 11% of total SX scores (step 1 of the regression model). The addition of sex, academic degree, age, and ethnicity into the regression model did not significantly increase predictive power (step 2 of the regression model).

The total FQR score (FQR1 + FQR2) also was examined as a predictor for the total SX score while accounting for sex, academic degree, age, and ethnicity (Table 6). This regression analysis revealed that the total FQR score was a better predictor than the total FQ score (compare R^2 values, step 1 of Tables 5 and 6). The total FQR score accounted for 52% of the total SX score (step 1 of Table 6). Again, the addition of sex, academic degree, age, and ethnicity did not significantly increase the predictive power of the model (step 2 of Table 6). In both regression models, R^2 shrinkage (the difference between R^2 and adjusted R^2 values) was less than 5% for the initial regression step, suggesting good generalizability for these simple linear models. The relationship between total FQR and total SX scores is plotted in Figure 1.

DISCUSSION

The primary purpose of this study was to determine if the use of formative quizzes and SGQQs would improve performance on subsequent summative exams. On first consideration, the putative benefit of formative quizzes and student participation in writing quiz questions might go unquestioned. Researchers have reported that formative assessments enhance summative exam performance in dental students,¹² medical students,¹⁵ and a variety of

Table 3 - Assessment Scores

Classes ^a	Class 1 Mean (95% CI)	Class 2 Mean (95% CI)	Class 3 – Control Mean (95% CI)
Sample size	65	48	76
FQ total %	75.73 (73.79–77.67)	80.12 (77.98–82.26)	–
FQR total %	90.43 (88.87–92.00)	92.73 (91.06–94.41)	–
SX total %	84.50 (82.70–86.30)	85.73 (83.11–88.35)	80.54 (78.38–82.71)

^a Class 1 students wrote and received formative quizzes. Class 2 students received, but did not write, formative quizzes. Class 3 (control) students neither wrote nor received formative quizzes.

Table 4 - One-Way ANOVA Assessing Summative Exam Score Differences

	df	Sums of Squares	Mean Square	F	P, 2-Tailed
Source of variation					
Between groups	2	953.44	476.72	6.37	.002
Within groups, error	186	13931.15	74.90	—	—
Total	188	14884.59			
	df	SE	r	t-Test	P, 2-Tailed
Planned contrasts					
Classes 1 & 2 vs class 3	186	2.58	.25	3.544	.000
Class 1 vs class 2	186	1.65	.05	.746	.456

Groups = (Class 1, Class 2, Class 3); Outcome = Total Summative Exam Score; $\omega^2 = .054$.

undergraduate majors.¹⁴ However, other investigators have reported that formative assessments do not enhance summative exam scores.^{7,13,17}

Our study results supported the notion that formative quizzes will improve performance on subsequent written summative exams. Both classes receiving formative quizzes (Classes 1 and 2) had significantly higher summative exam scores than Class 3, the class not receiving these quizzes.

We also anticipated that writing SGQQs would require greater study and understanding, and this would produce an additional increase in summative exam scores. This argument is consistent with the theory that instructional methods that promote learner interaction are more effective than less active methods.¹⁸ However, this hypothesis was not supported by our findings. The class that produced SGQQs and also received the formative quizzes (Class 1) did not have significantly higher summative exam scores than Class 2, which received only the formative quizzes.

Several studies have suggested that formative assessments can predict summative exam outcomes.^{4,9,11} Our study results supported this conclusion. Total FQ scores and total FQR scores were significant predictors of written summative exam scores. However, total FQR scores were substantially stronger predictors than total FQ scores. With either FQs or FQRs, sex, age, academic degrees, and ethnicity were not significant copredictors of summative exam scores.

Table 5 - Multiple Regression for Formative Quizzes Predicting Summative Exam Scores

	B	SE B	β	P, 2-Tailed
Step 1				
Constant	58.91	7.09		.000
Formative quiz totals	0.34	0.09	.33	.000
Step 2				
Constant	64.12	8.23		.000
Formative quiz totals	0.32	0.09	.33	.001
Sex	-1.52	1.49	-.09	.307
Age	-1.91	2.22	-.08	.392
Ethnicity	-1.15	1.81	-.06	.525
Academic degree	1.57	2.66	-.06	.556

Step 1, $R^2 = .11$ ($P = .000$); Step 2, $\Delta R^2 = .02$ ($P = .641$).

In a recent comparative review, Dunlosky et al¹⁹ explored the efficacy of 10 learning techniques: elaborative interrogation, self-explanation, summarization, highlighting/underlining, keyword mnemonic, imagery for text, rereading, practice testing, distributed practice, and interleaved practice. They reported that practice testing, which they defined as “self-testing or taking practice tests over to-be-learned material,” has demonstrated effects across an impressive range of practice-test formats, kinds of material, learner ages, outcome measures, and retention intervals. Moreover, they noted that practice testing is not particularly time intensive relative to the other learning techniques, and it can be implemented with minimal training.

Dunlosky et al¹⁹ emphasized the importance of instructor feedback in association with practice testing. They noted that instructor feedback protects against perseveration errors when students respond incorrectly on a practice test. In addition, they commented that the corrective effect of feedback does not require that it be presented immediately after the practice test. In fact, Metcalfe et al²⁰ found that final-test performance for initially incorrect responses was actually better when feedback had been delayed than when it had been immediate.

Finally, it is impressive that the beneficial effects of practice testing have been observed for substantial time periods after the exercise: 2–4 weeks,^{21–25} 2–4 months,^{26–28}

Table 6 - Multiple Regression for Formative Quiz Reviews Predicting Summative Exam Scores

	B	SE B	β	P, 2-Tailed
Step 1				
Constant	-0.90	7.85		.909
Formative quiz totals	0.94	0.09	.72	.000
Step 2				
Constant	2.37	8.63		.784
Formative quiz totals	0.94	0.09	.72	.000
Sex	-1.47	1.09	-.09	.181
Age	-0.05	1.64	-.00	.974
Ethnicity	-0.31	1.32	-.02	.813
Academic degree	-0.64	1.96	-.02	.744

Step 1, $R^2 = .52$ ($P = .000$); Step 2, $\Delta R^2 = .01$ ($P = .659$).

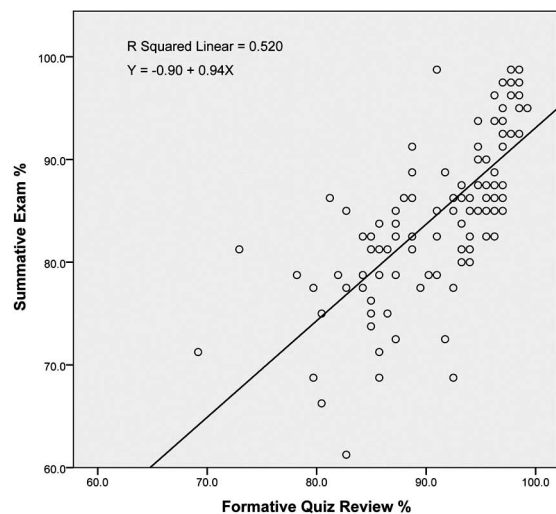


Figure 1 - Formative quiz review predicting summative exam performance. Total formative review score (FQR1 + FQR2) was a good predictor of total summative exam performance, predicting 52% of that performance ($R^2 = .52$). Specifically, our data suggested that a 1-unit change in FQ total score is associated with a 0.94-unit change in the total summative exam score ($B = 0.94$, also see Table 6).

5–8 months,^{29,30} 9–11 months,³¹ and even 1–5 years.³² These are exciting findings for students and educators. We seek long-lasting knowledge, not just temporary learning improvements.

Study limitations are that students in the experimental group received formative quizzes (FQ₁₋₄ and FQ₅₋₈) and quiz reviews (FQR1 and FQR2) before the 2 summative session exams (SX1 and SX2). Students in the control group received traditional lectures on the same topic material and were assessed with similar summative session exams, but did not receive the formative quizzes or quiz reviews. The current study design did not allow us to parse the relative effect of formative quizzes from that of the quiz reviews. Would FQs without FQRs be effective? A future study with separate experimental groups for each of these factor combinations is needed to make that determination. In addition, another limitation of this study is the relatively small sample size and restricted source. Ours was a “sample of convenience” as all students came from a single chiropractic college. Future studies are needed that examine larger populations drawn from a representative sampling of chiropractic colleges. One also might wonder if the observed effect would be substantially influenced by topic of study. All students in this study were enrolled in a physical examination course, and only written assessments are reported here. Other researchers have found that “practice testing” is robust, producing beneficial affects across many topics.¹⁹

CONCLUSION

It is reasonable to posit that the formative quizzes in this study enhanced and predicted summative written exam scores because these quizzes were similar in form and

content to the written summative exams and they evaluated the same knowledge base. We concluded that nonthreatening formative quizzes, with faculty feedback and quiz reviews, may be used to enhance student learning and suggested that they also may serve to identify students who, without additional remediation, will perform poorly on subsequent summative written exams. Moreover, an extensive education literature suggests that the beneficial effects of active learning may be quite durable, lasting months or even years.

FUNDING AND CONFLICTS OF INTEREST

This work was funded internally. The authors declare that there are no conflicts of interest to declare relevant to this work.

About the Authors

Niu Zhang is an assistant professor at Palmer College of Chiropractic Florida (4777 City Center Parkway, Port Orange, FL 32129; niu.zhang@palmer.edu). Charles Henderson is a consultant with Henderson Technical Consulting (5961 Broken Bow Lane, Port Orange, FL 32127; cnrhenderson@gmail.com). Address correspondence to Niu Zhang, 4777 City Center Parkway, Port Orange, FL 32129; niu.zhang@palmer.edu. This article was received April 10, 2014, revised September 11, 2014, and accepted September 13, 2014.

© 2015 Association of Chiropractic Colleges

REFERENCES

1. Beecher ME, Fisher L. High school course and scores as predictors of college success. *J Coll Admiss.* 1999; 163:4–9.
2. Collier VP. A synthesis of studies examining long-term language minority student data on academic achievement. *Biling Res J.* 1992;16(1–2):187–212.
3. DeBerard MS, Spielmans GI, Julka DL. Predictors of academic achievement and retention among college freshmen: a longitudinal study. *Coll Stud J.* 2004;38: 66–80.
4. Fowell SL, Southgate LJ, Bligh JG. Evaluating assessment: the missing link? *Med Educ.* 1999;33(4): 276–281.
5. Hattie JA. Identifying the salient factors of a model of student learning: synthesis of meta-analyses. *Int J Educ Res.* 1987;4:187–212.
6. Seale JK, Chapman J, Davey C. The influence of assessments on students’ motivation to learn in a therapy degree course. *Med Educ.* 2000;34(8):614–621.
7. Haberyan KA. Do weekly quizzes improve student performance on general biology exams? *Am Biol Teach.* 2003;65:110–114.

8. Rolfe I, McPherson J. Formative assessment: how am I doing? *Lancet*. 1995;345(8953):837–839.
9. Padilla-Walker LM. The impact of daily extra credit quizzes on exam performance. *Teach Psychol*. 2006;33: 236–239.
10. Leaf DE, Leo J, Smith PR, et al. SOMOSAT: Utility of a web-based self-assessment tool in undergraduate medical education. *Med Teach*. 2009;31(5):e211–e219.
11. Kibble J. Use of unsupervised online quizzes as formative assessment in a medical physiology course: effects of incentives on student participation and performance. *Adv Physiol Educ*. 2007;31(3):253–260.
12. Olson BL, McDonald JL. Influence of online formative assessment upon student learning in biomedical science courses. *J Dent Educ*. 2004;68(6):656–659.
13. Brothen T, Wambach C. Effective student use of computerized quizzes. *Teach Psychol*. 2001;28:292–294.
14. Buchanan T. The efficacy of a World-Wide Web mediated formative assessment. *J Comput Assist Learn*. 2000;16:193–200.
15. Brar MK, Laube DW, Bett GC. Effect of quantitative feedback on student performance on the National Board Medical Examination in an obstetrics and gynecology clerkship. *Am J Obstet Gynecol*. 2007; 197(5):530–535.
16. Kirk RE. Practical significance: a concept whose time has come. *Educ Psychol Meas*. 1996;56(5):746–759.
17. Peat M, Franklin S. Has student learning been improved by the use of online and offline formative assessment opportunities? *Aust J Ed Technol*. 2003; 19(1):87–99.
18. Cook DA, Thompson WG, Thomas KG, Thomas MR, Pankratz VS. Impact of self-assessment questions and learning styles in Web-based learning: a randomized, controlled, crossover trial. *Acad Med*. 2006;81(3): 231–238.
19. Dunlosky J, Rawson KA, Marsh EJ, Nathan MJ, Willingham DT. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychol Sci Public Interest*. 2013;14(1):4–58.
20. Metcalfe J, Kornell N, Son LK. A cognitive-science based programme to enhance study efficacy in a high and low risk setting. *Eur J Cogn Psychol*. 2007;19(4–5): 743–768.
21. 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.
22. Butler AC, Roediger HL. Testing improves long-term retention in a simulated classroom setting. *Eur J Cogn Psychol*. 2007;19(4–5):514–527.
23. Carpenter SK, Pashler H, Wixted JT, Vul E. The effects of tests on learning and forgetting. *Mem Cognit*. 2008;36(2):438–448.
24. Kromann CB, Jensen ML, Ringsted C. The effect of testing on skills learning. *Med Educ*. 2009;43(1):21–27.
25. Rohrer D, Taylor K, Sholar B. Tests enhance the transfer of learning. *J Exp Psychol Learn Mem Cogn*. 2010;36(1):233–239.
26. McDaniel MA, Anderson JL, Derbish MH, Morrisette N. Testing the testing effect in the classroom. *Eur J Cogn Psychol*. 2007;19(4–5):494–513.
27. Morris PE, Fritz CO. The improved name game: better use of expanding retrieval practice. *Memory*. 2002; 10(4):259–266.
28. Rawson KA, Dunlosky J. Optimizing schedules of retrieval practice for durable and efficient learning: how much is enough? *J Exp Psychol Gen*. 2011;140(3): 283–302.
29. McDaniel MA, Agarwal PK, Huelser BJ, McDermott KB, Roediger HL. Test-enhanced learning in a middle school science classroom: the effects of quiz frequency and placement. *J Educ Psychol*. 2011;103(2):399–414.
30. Kromann CB, Bohnstedt C, Jensen ML, Ringsted C. The testing effect on skills learning might last 6 months. *Adv Health Sci Educ Theory Pract*. 2010;15(3):395–401.
31. Carpenter SK, Pashler H, Cepeda NJ. Using tests to enhance 8th grade students' retention of U.S. history facts. *Appl Cogn Psychol*. 2009;23(6):760–771.
32. Bahrick HP, Bahrick LE, Bahrick AS, Bahrick PE. Maintenance of foreign language vocabulary and the spacing effect. *Psychol Sci*. 1993;4(5):316–321.