

Integrating Cognitive Science and Technology Improves Learning in a STEM Classroom

Andrew C. Butler · Elizabeth J. Marsh · J. P. Slavinsky ·
Richard G. Baraniuk

Published online: 18 March 2014

© Springer Science+Business Media New York 2014

Abstract The most effective educational interventions often face significant barriers to widespread implementation because they are highly specific, resource intense, and/or comprehensive. We argue for an alternative approach to improving education: leveraging technology and cognitive science to develop interventions that generalize, scale, and can be easily implemented within any curriculum. In a classroom experiment, we investigated whether three simple, but powerful principles from cognitive science could be combined to improve learning. Although implementation of these principles only required a few small changes to standard practice in a college engineering course, it significantly increased student performance on exams. Our findings highlight the potential for developing inexpensive, yet effective educational interventions that can be implemented worldwide.

Keywords Education · Technology · Retrieval practice · Spacing · Feedback · Transfer of learning

Introduction

Improving education is an immense and complex global challenge. Millions of students worldwide face significant barriers to obtaining education, particularly those who come from disadvantaged backgrounds (OECD 2012). Even with access to education, too many primary and secondary students consistently score below grade-specific proficiency levels in math and science on international assessments (e.g., Trends in International Mathematics and Science Study, Program for International Student Assessment); furthermore, students who demonstrate proficiency in science and mathematics often struggle to apply their knowledge (Mullis et al. 2012; OECD 2010). In the subset of students who reach higher education, there are also signs of trouble: post-secondary graduates often lack the advanced understanding and skills needed to compete in a modern knowledge-based economy (U.S. National Science Board 2012; UNESCO 2011). With tuition growing faster than household incomes and school

A. C. Butler (✉) · E. J. Marsh
Department of Psychology & Neuroscience, Duke University, Box 90086, Durham, NC 27708-0086, USA
e-mail: andrew.butler@duke.edu

J. P. Slavinsky · R. G. Baraniuk
Department of Electrical and Computer Engineering, Rice University, Houston, USA

budgets facing severe cuts, we need to make educational interventions easier and cheaper to apply.

In a concerted effort to address this crisis, educators and scientists from a variety of disciplines have produced numerous highly effective interventions, ranging from new instructional methods to computer-based learning systems. However, there are significant barriers to implementing these interventions across an entire education system, which diminishes their potential impact. For example, many of these interventions involve a complete overhaul of existing curricula and pedagogy (e.g., Michaelsen et al. 2002); require enormous investments of time, money, and expertise (e.g., Murray 1999); target a highly specific set of knowledge (e.g., Anderson et al. 1985); and/or necessitate considerable efforts on the behalf of overworked educators (e.g., Armbruster et al. 1987; Kirschner et al. 2006).

Although there is certainly merit in interventions that are highly specific, resource-intense, and comprehensive, we argue that a different approach may be more fruitful. The goal of our approach is to develop interventions that generalize, scale, and can be easily implemented within any curriculum while minimizing any loss in effectiveness. In accordance with the National Education Technology Plan of the United States (U.S. Department of Education 2010), we believe that the key to developing such interventions is leveraging both rapidly advancing technology and research from cognitive science (Bransford et al. 1999) to put robust learning tools in the hands of educators and students. The integration of simple, yet powerful principles of learning into advanced technologies creates the potential to apply effective practices to education systems worldwide.

We report the results of a classroom experiment that demonstrates the promise of this approach to improving education. Previous studies have achieved large positive effects on learning by completely changing the curriculum and pedagogy (e.g., Deslauriers et al. 2011) or using carefully controlled laboratory conditions and simple materials (e.g., Karpicke and Roediger 2008). In contrast, our experiment investigated whether a few small, but important changes to standard practice could make a big difference in a “noisy” and complicated classroom setting. In making these changes, we deliberately combined multiple principles of learning in order to increase the effectiveness of the intervention. We examined whether this simple, but potentially powerful intervention would benefit student learning in spite of a host of uncontrolled factors that should diminish its effectiveness.

The experiment was carried out in the fall semester of 2012 in an upper-level class (“Signals and Systems”) that is a core requirement for undergraduate electrical and computer engineering majors at Rice University. The course deals with signals, systems, and transforms, from their theoretical mathematical foundations to practical implementation in circuits and computer algorithms. Students attended 3 h of lecture per week and viewed texts, videos, and simulations outside the classroom. Grading in the course was based on two exams (a midterm and final), weekly homework assignments, a group project, and class participation. The intervention targeted the weekly homework assignments and implemented three principles from cognitive science that are robust, highly replicable, and generalizable: repeated retrieval practice, spacing, and feedback (see Table 1). These principles have been shown to increase long-term retention and promote transfer of learning by fostering deeper understanding. Based on the wealth of evidence supporting their effectiveness, all three principles have been recommended by the US Institute of Education Sciences (Pashler et al. 2007). In the basic method used for the intervention, students were provided with repeated opportunities to retrieve and use their knowledge through a schedule of spaced practice; they were also required to view timely feedback in order to reinforce their knowledge and to correct their errors. The main question of interest was whether these relatively minor changes would

improve exam performance compared to standard practice, even though the changes only affected the weekly homework assignments.

Methods

Participants

Forty Rice University undergraduate students participated in the experiment (four additional students in the course chose not to release their data for research purposes). Each student who participated received a \$25 Amazon.com gift certificate.

Design and Counterbalancing

In order to minimize disruption to the way the class was normally taught, the intervention was only implemented for the weekly homework assignments. The intervention was delivered through the use of OpenStax Tutor (<http://openstaxtutor.org>), a new e-learning system designed to implement cognitive science principles. We assessed the effectiveness of the intervention by comparing it to standard practice, which mirrored the way homework assignments were administered in previous semesters. In both cases, students worked the homework problems, turned in solutions by the deadline, and then received feedback (see Fig. 1 for an example of a typical problem used in the course). Learning under the intervention differed from standard practice in four small, but important ways:

Table 1 Descriptions of three principles from cognitive science that increase long-term retention and promote transfer of learning as well as how each principle was implemented in the intervention in contrast to standard practice

Principle	Description	Intervention	Standard practice
Repeated retrieval practice	Repeatedly retrieving information from memory strengthens memory for that information; it can also improve understanding of that information (Roediger and Butler 2011)	Students solved three sets of problems on each topic	Students solved only a single set of problems for each topic
Spacing	Spacing or distributing practice over time produces a better long-term retention than massing practice (i.e., cramming) (Cepeda et al. 2006)	Practice on the three sets of problems was distributed over 3 weeks	After material was covered in the course, it was not revisited on the homework assignments
Feedback	Feedback provides learners with information that enables them to correct errors and to improve understanding (Hattie and Timperley 2007). Immediate feedback is often more effective in the classroom (Kulik and Kulik 1988)	Feedback was accessible immediately after the assignment deadline; students were required to view the solution to each problem in order to receive credit for completing the assignment	Feedback was accessible 1 week after the assignment deadline; students were not required to view it

1. *Repeated Retrieval Practice.* Under standard practice, students received one set of problems related to the material covered in class during a given week (i.e., a single opportunity to practice retrieving and using their knowledge). Under the intervention, students received the same set of initial problems, plus another two sets of follow-up problems (i.e., three opportunities to practice retrieving and using their knowledge). Students were told that their performance on the extra follow-up problems would only count towards their participation grade in the class.
2. *Spacing.* As in most college courses, standard practice meant that once material was covered in class and on a corresponding homework assignment, it was not revisited until the exam. In contrast, the three problem sets that students received in the intervention were spaced out by giving the follow-up practice problems on the next two homework assignments (e.g., problems on the material covered in week 1 were given on assignments #1, #2, and #3).
3. *Timely Feedback.* Standard practice provided feedback 1 week after the deadline (i.e., a common delay in education that allows time for teaching assistants to grade the assignment). The intervention delivered feedback immediately after the assignment deadline.
4. *Required Feedback Viewing.* Standard practice meant that feedback viewing was optional, as instructors generally have no control or knowledge of whether their students process the feedback. However, the intervention required feedback viewing; students did not receive credit for completing each assignment until they had viewed the feedback.

Overall, the intervention provided students with spaced opportunities to repeatedly practice retrieving and using their knowledge while also encouraging them to fully process the feedback.

The experiment utilized a single factor (intervention vs. standard practice), within-subjects design. At the start of the semester, students were randomly assigned to one of two groups. Both groups alternated across weeks between the intervention and standard practice, with the only difference being that group 1 began with the intervention in week 1, whereas group 2 began with standard practice. Thus, in any given week, one half of the students were assigned to the intervention, while the other half received standard practice. Table 2 provides a schematic representation of how the experiment was implemented within the course on a week-by-week basis. The use of a within-subjects experimental design alleviated any concerns about individual differences in ability.

Materials

The class covered 11 topics (e.g., discrete Fourier transforms; see Table 2 for a partial list) for which there was a corresponding homework assignment. Each topic contained approximately five core concepts (e.g., circular convolution, delta sifting property, etc.), which were generally taught over the course of a week. The topics were designed to be relatively independent of each other in order to minimize the potential for learning from the intervention to affect learning from standard practice (or vice versa), as students switched back and forth between the two methods from week to week; however, as in any course, the topics were somewhat related and built upon each other to a degree.

Three practice problems were created for each concept. Each practice problem was unique and required the application of the concept to determine the solution (see Fig. 1). For a given topic, each of the three problems related to a concept was assigned to one of three problem sets. One problem set was always used as the initial problems for that topic, while the other

Table 2 Schematic representation of how the experiment was implemented during the first half of the course. The same pattern was continued during the second half of the course, which concluded with the final exam

Week	Topic	Topic code	Group 1 assignment		Topic code	Group 2 assignment	
			Initial problems (method)	Follow-up problems		Initial problems (method)	Follow-up problems
1	Signal basics	A	Topic A (intervention)	–	A	Topic A (standard practice)	–
2	Time-domain analysis of continuous-time systems	B	Topic B (standard practice)	Topic A	B	Topic B (intervention)	–
3	Time-domain analysis of discrete-time systems	C	Topic C (intervention)	Topic A	C	Topic C (standard practice)	Topic B
4	Continuous-time Fourier series	D	Topic D (standard practice)	Topic C	D	Topic D (intervention)	Topic B
5	Discrete Fourier transform	E	Topic E (intervention)	Topic C	E	Topic E (standard practice)	Topic D
6	Fast Fourier transform and orthonormal bases	F	Topic F (standard practice)	Topic E	F	Topic F (intervention)	Topic D
7	Midterm exam		Topics A, B, C, and D ^a			Topics A, B, C, and D ^a	

^a Only the topics from the first 4 weeks were tested on the midterm exam

two problem sets were used to provide additional practice in subsequent weeks for the intervention (i.e., the follow-up problems). The order of the problem sets was not counterbalanced for ethical purposes; every student in the class received the same initial problems because these problems were fully graded and counted toward the homework grade in the class. The follow-up practice problems were graded pass/fail and only counted towards the participation grade.

Another set of 34 problems was created for the midterm and final exams. Like the practice problems, each exam problem was unique and required the application of a concept. Due to time limitations, the exams did not contain a problem for every concept. However, every topic was tested and received equal coverage on the exam (approximately three problems per topic). Topics 1–4 were tested on the midterm exam, and topics 5–11 were tested on the final exam.

Procedure

At the beginning of the semester, a researcher visited the classroom and explained to students that they would have the option of participating in a study. They were told that the long-term goal of the research project was to create a personalized learning system for educators and students. They were also told that the research team was conducting experiments to investigate questions about how people learn and that their interactions with the OpenStax Tutor system were of interest for research purposes. Students were not informed about the specifics of the manipulation, counterbalancing (e.g., students receiving the intervention and standard practice on different topics), or the hypotheses of the experiment in order to minimize demand characteristics. When they first logged into the OpenStax Tutor system, they had the opportunity to consent electronically and could change their decision at any time up until the end of the course.

Problem

Consider an LTI system with impulse response

$$h[n] = \delta[n] - \delta[n - 1].$$

Calculate $H(e^{j\omega})$, the DTFT of $h[n]$. Simplify it as much as you can (in this case, so that there are no sums in the answer).

- a) $H(e^{j\omega}) = 2e^{-j\omega/2} \cos(\omega/2)$
- b) $H(e^{j\omega}) = 2e^{j\omega/2} \cos(\omega/2)$
- c) $H(e^{j\omega}) = e^{\omega} \cos(\frac{\pi}{2} \omega - \pi/4)$
- d) $H(e^{j\omega}) = e^{\omega} \cos(\frac{\pi}{2} \omega + \pi/4)$
- e) $H(e^{j\omega}) = 2 \cos(\omega/2) e^{j(-\omega/2 + \pi/2)}$
- ✓ f) $H(e^{j\omega}) = 2 \sin(\omega/2) e^{j(-\omega/2 + \pi/2)}$
- g) $H(e^{j\omega}) = e^{\omega n} \cos(\frac{\pi}{2} n)$
- h) $H(e^{j\omega}) = e^{\omega n} \sin(\frac{\pi}{2} n)$

Feedback

High-Level Explanation

Calculate the DTFT using the formula, factoring out an exponential to simplify the expression.

Detailed Solution

$$\begin{aligned}
 H(e^{j\omega}) &= \sum_{n=-\infty}^{\infty} h[n] e^{-j\omega n} \\
 &= \sum_{n=-\infty}^{\infty} (\delta[n] - \delta[n - 1]) e^{-j\omega n} \\
 &= 1 e^{-j\omega 0} - 1 e^{-j\omega 1} \\
 &= 1 - e^{-j\omega} \\
 &= e^{-j\omega/2} (e^{j\omega/2} - e^{-j\omega/2}) \\
 &= 2j \sin(\omega/2) e^{-j\omega/2} \\
 &= 2 \sin(\omega/2) e^{j(-\omega/2 + \pi/2)}
 \end{aligned}$$

Fig. 1 A sample problem from the experiment, which involves computing a discrete-time Fourier transform. At first, the problem was presented in OpenStax Tutor without any of the multiple-choice alternatives. After the students worked the problem and entered a free-form response, they were presented with the alternatives and selected a response. Finally, they received feedback in the form of a high-level explanation and a detailed solution

In order to mitigate any novelty effects (Clark 1983), all students used the OpenStax Tutor system to enter their solutions to the homework problems and to view feedback regardless of whether the material was assigned to the intervention condition or standard practice. Each assignment contained approximately 10–20 problems, including a set of initial problems on the concepts covered that week and follow-up problems to provide additional retrieval practice on material from previous weeks.

Answering each problem involved two steps: (1) entering a response in a free-form text box and then (2) selecting the correct response from a set of multiple-choice alternatives, which appeared after the free-form response was submitted. This two-step answering process was designed to maximize the mnemonic benefit of retrieval practice while retaining the objective and automated scoring of the multiple-choice format. For all homework assignments, students were allowed to work the problems in groups (which is standard practice in the course), but they were required to input their final answers individually as well as to view the feedback individually. Feedback indicated the correct multiple-choice alternative and provided a solution to the problem.

Two take-home exams (a midterm and a final) were used to assess learning of the material covered in the homework assignments. Both exams were administered through the OpenStax Tutor system and consisted of new problems that required students to apply their knowledge of the core concepts from the course. Students received 1 week to complete each exam, and they entered their solutions to the problems using the same two-step answering procedure as the homework assignments. The exams were closed book, and students were required to work the problems individually under the Rice University honor code. After the semester, the data from the fully graded initial problems, the follow-up problems, and exams were analyzed for students who consented to release their data for research purposes.

Results

The results of the experiment were clear; the combination of small, but important changes to a small part of standard practice boosted student learning and retention in the course. Students performed better on the exam problems about material learned via the intervention than they did on problems about material learned through standard practice. Given that a two-step answering process was used (free form followed by multiple choice), it was important to confirm that the result held for both question formats. It did. The advantage in performance was evident in both the free-form responses [60 vs. 53 %; $t(39)=2.28$, $p=0.03$, $d=0.34$] as well as the subsequent answers selected from among multiple-choice alternatives [76 vs. 69 %; $t(39)=2.98$, $p=0.01$, $d=0.34$].

These differences in exam performance were not due to differences in learning that occurred before the homework assignments or the inherent difficulty of the course material; students correctly answered the majority of the initial set of homework problems, and performance did not differ as a function of the intervention versus standard practice (90 vs. 90 %; $t<1$). Another possibility is that students interpreted the follow-up problems as signaling that the intervention topics were the most important ones, leading them to focus more on these topics when preparing for the exams. However, such targeted studying was unlikely, given that students were told that all the topics would be covered on the exams and the practice problems only counted toward students' participation grades. Moreover, all topics received equal coverage on the exams, and so presumably, any students adopting such a strategy would have noted this fact after the midterm and changed their strategy. If that were the case, then one

would expect the overall effect to be driven by differences in performance on the midterm exam; however, when the midterm exam performance was excluded, there was still a significant difference between the intervention and standard practice on the final exam [short answer, 63 vs. 56 %, $t(39)=2.05$, $p=0.047$, $d=0.36$, and multiple choice, 72 vs. 67 %, $t(39)=2.59$, $p=0.013$, $d=0.22$]. Yet, another possibility is that retrieval-induced forgetting (Anderson et al. 1994) played a role in producing the differences in exam performance. However, this explanation is unlikely because several of the boundary conditions under which retrieval-induced forgetting effects disappear are present in our study: relational processing of the material during learning (e.g., Anderson and McCulloch 1999), the use of specific retrieval cues on the criterial test (i.e., as opposed to category cues; e.g., Butler et al. 2001), and retention intervals longer than 24 h (e.g., MacLeod and Macrae 2001).

Finally, as expected, the experimental manipulation influenced students' feedback viewing behavior. On average, they viewed feedback for each problem a significantly greater number of times under the intervention relative to standard practice [2.1 vs. 1.3; $t(39)=10.55$, $p=0.001$, $d=0.97$]. In addition, students completely failed to view feedback 30 % of the time when standard practice was used, but only 4 % of the time under the intervention.

Discussion

Our small-to-medium effect size is impressive when considered within the broader context (Hattie 2009). The experiment was conducted in the classroom, and no attempts were made to control numerous factors that would be expected to reduce or eliminate any effect of the intervention. In contrast with laboratory studies where the learning process is highly controlled and the potential for outside learning is extremely low (e.g., Karpicke and Roediger 2008), students who participated in the present experiment had a multitude of opportunities to learn the material beyond the homework assignments that delivered the intervention (e.g., attending lectures, reading texts, watching videos, using simulations, studying with classmates, etc.). One approach to classroom research is to minimize the impact of such extra-experimental learning by either exerting complete control over the curriculum and pedagogy or “roping off” material so that students do not engage with it outside of the experimental aspect of the course. We took the opposite approach by giving students unrestricted access to the materials and allowing them to control their own learning (aside from the homework assignments). Moreover, it is important to note that the students were taking the course for credit, so they were motivated to excel on the exams and likely prepared for them in many ways beyond the homework assignments (e.g., attending review sessions, meeting with a teaching assistant, etc.).

The effect of our intervention in this noisy classroom environment is even more impressive when compared to other educational interventions in terms of cost, generalizability, and the degree of disruption to standard practice. Our intervention was highly cost-effective and could easily be applied to other courses in a host of disciplines. On average, intelligent tutoring systems produce larger effects on learning ($d=0.76$) when compared with no tutoring (VanLehn 2011); however, these interventions are enormously expensive to develop in terms of money, time, and resources, and they target a highly specific set of knowledge (e.g., algebra). The changes implemented in our intervention caused a minimal disruption to the class and required only a modest amount of work by the instructor. Nevertheless, our intervention produced a larger effect than the average for interventions that completely overhaul curricula and/or pedagogy, such as comprehensive teaching reforms ($d=0.22$) or instructional methods like problem-based learning ($d=0.15$) (Hattie 2009).

Our findings provide compelling evidence that combining the principles of repeated retrieval practice, spacing, and feedback generalizes to higher-order learning. The material covered in the Signals and Systems course is far more complex than the types of materials commonly used in basic laboratory research, which are often fact-based and verbal in nature (for discussion, see Dunlosky et al. 2013). Prior research in the classroom has demonstrated the efficacy of these principles individually (e.g., Carpenter et al. 2009; McDaniel et al. 2011), but these studies tend to use simple fact-based materials as well. Furthermore, the vast majority of studies on these principles, both in the laboratory and the classroom, have focused on the retention of knowledge (e.g., using the same or very similar questions during initial retrieval practice and on the final test). In contrast, the present experiment showed that these principles promote the acquisition of knowledge that transfers to different contexts as measured by the ability to solve new application problems on the exams.

In summary, the combination of spaced retrieval practice and required feedback viewing had a powerful effect on student learning of complex engineering material. Of course, the principles from cognitive science could have been applied without the use of technology. However, our belief is that advances in technology and ideas from machine learning have the potential to exponentially increase the effectiveness and impact of these principles. Automation is an important benefit, but technology also can provide a personalized learning experience for a rapidly growing, diverse body of students who have different knowledge and academic backgrounds. Through the use of data mining, algorithms, and experimentation, technology can help us understand how best to implement these principles for individual learners while also producing new discoveries about how people learn. Finally, technology facilitates access. Even if an intervention has a small effect size, it can still have a substantial impact if broadly implemented. For example, aspirin has a small effect on preventing heart attacks and strokes when taken regularly, but its impact is large because it is cheap and widely available. The synergy of cognitive science, machine learning, and technology has the potential to produce inexpensive, but powerful learning tools that generalize, scale, and can be easily implemented worldwide.

Acknowledgments The authors would like to thank Daniel Williamson, Matthew Moravec, Eva Dyer, Kevin Burleigh, and Kim Davenport for their contributions to this research. This research was supported by NSF grant no. IIS-1123617 to EJM and NSF grant no. IIS-1124535 and Google Faculty Research Award to RGB.

All authors contributed to the idea for the research. ACB and EJM designed the experiment. JPS directed the creation and implementation of the software infrastructure for OpenStax Tutor. RGB assisted in the design of OpenStax Tutor and taught the course. ACB analyzed the data and drafted the manuscript. All authors edited the manuscript. Correspondence and requests for materials and data should be addressed to ACB (andrew.butler@duke.edu).

References

- Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. *Science*, 228, 456–462.
- Anderson, M. C., & McCulloch, K. C. (1999). Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 608–629.
- Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1063–1087.
- Armbruster, B. B., Anderson, T. H., & Ostertag, J. (1987). Does text structure/summarization instruction facilitate learning from expository text? *Reading Research Quarterly*, 22, 331–346.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: brain, mind, experience and school*. Washington, D.C.: National Academy.
- Butler, K. M., Williams, C. C., Zacks, R. T., & Maki, R. H. (2001). A limit on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1314–1319.

- Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology*, *23*, 760–771.
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychological Bulletin*, *132*, 354–380.
- Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, *53*, 445–459.
- Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science*, *332*, 862–864.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*, 4–58.
- Hattie, J. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112.
- Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science*, *319*, 966–968.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*, 75–86.
- Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58*, 79–97.
- MacLeod, M. D., & Macrae, C. N. (2001). Gone but not forgotten: the transient nature of retrieval-induced forgetting. *Psychological Science*, *12*, 148–152.
- McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L. (2011). Test-enhanced learning in a middle school science classroom: the effects of quiz frequency and placement. *Journal of Educational Psychology*, *103*, 399–414.
- Michaelsen, L. K., Knight, A. B., & Fink, L. D. (Eds.). (2002). *Team-based learning: a transformative use of small groups*. Westport: Praeger.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in science and mathematics*. Chestnut Hill: TIMSS & PIRLS International Study Center, Boston College.
- Murray, T. (1999). Authoring intelligent tutoring systems: an analysis of the state of the art. *International Journal of Artificial Intelligence in Education*, *10*, 98–129.
- OECD. (2010). *PISA 2009 results: what students know and can do—student performance in reading, mathematics and science (volume I)*. Paris: OECD.
- OECD. (2012). *Education at a glance 2012: OECD indicators*. Paris: OECD.
- Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., et al. (2007). *Organizing instruction and study to improve student learning: a practice guide (NCER 2007–2004)*. Washington, D.C.: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.
- Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*, 20–27.
- U.S. Department of Education. (2010). *Transforming American education: learning powered by technology*. Washington DC: Office of Educational Technology, National Education Technology Plan 2010.
- U.S. National Science Board. (2012). *Science and engineering indicators 2012*. Arlington: National Science Foundation (NSB 12–01).
- UNESCO Institute for Statistics (UIS). (2011). *Global education digest 2011*. Montreal: UIS.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*, 197–221.