Title

Improving Long-term Retention of Mathematical Knowledge through Automatic Reassessment and Relearning

Abstract

This study investigates whether an online tutor that automatically reassesses student knowledge and provides relearning opportunities improves long-term retention for individual students and across specific mathematical skills. Over a single school year, 97 8th-grade students were pretested on a set of mathematical skills and then subsequently practiced them until mastery was attained on all skills. After mastery, students received automatic online reassessment and relearning opportunities for half the skills, while the other half were not further assessed. End-of-year post-test results revealed better performance on the problem sets that were reassessed and relearned versus not. Furthermore, reassessment and relearning was especially beneficial for students with low pre-test performance and for more difficult skills. The educational implications will be discussed.

Objectives

The current study is part of a larger project focused on applying two cognitively-based instructional design principles: (1) spaced practice and (2) focused quizzing with feedback, to the redesign of the *Connected Mathematics Program2* (CMP2) curriculum. One goal of the larger study is to conduct empirical studies of "optimal" patterns for the spacing of practice and feedback to inform the redesign of the CMP2 curriculum. The main objective of the current study is to investigate the benefits of two different practice-testing conditions across individuals and across mathematical contents and skills that are essential targets of the CMP2 curricula. We are interested in employing the principles of spacing and formative assessment to improve long-term retention of students' mathematical knowledge through an online *Automatic Reassessment and Relearning System* (ARRS). Specifically, we are examining (1) the degree of benefit of ARRS relative to a "baseline" mastery condition, (2) the relative benefit of ARRS for individuals who are more or less proficient in their mathematical skills and content, and (3) the relative value of ARRS for knowledge and skills that are more or less challenging.

Perspectives and theoretical framework

Decades of psychological and educational research have identified several beneficial design principles intended to increase long-term knowledge retention. Two principles in particular that have received a large amount of attention are the spacing of practice across time and formative assessment. Extensive research has demonstrated large retention advantages when learners are exposed to key facts, concepts, and knowledge at multiple points in time as opposed to one "massed" practice opportunity (see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). This phenomenon is dubbed the "spacing effect." Spaced practice is effective because it reinforces connections between ideas and provides multiple routes of retrieval for the same piece of information (Underwood, 1961). However, the empirical conclusions about spaced practice are often not utilized in teaching mathematical content. In a review of the spacing effect literature, Rohrer (2009) concluded, "the spacing of practice is being grossly underutilized in mathematics instruction." Other authors have arrived at the similar conclusions (Bahrick & Hall, 1991; Dempster, 1988; Mayfield & Chase, 2002; Willingham, 2002, Pashler, et al., 2007).

The spacing effect is complementary to a second design principle, formative assessment. Formative assessment can be defined as a process that provides feedback to inform and adjust ongoing learning to improve understanding (Black & Wiliam, 1998). A common use of formative assessment is quizzing; student knowledge of recent content is evaluated and feedback on student performance provides an informative tool to diagnose what students do not fully understand and what content requires further practice. Formative assessment has several advantages. First, periodic testing provides opportunities for students to practice retrieving knowledge and using skills and concepts. Testing knowledge is demonstrated to be more beneficial for knowledge retention than simply restudying (e.g. Butler & Roediger, 2007). This phenomenon is known as the "testing effect". Testing may be more successful compared to restudy in part because it prompts students to retrieve information actively, reflect on the state of their knowledge, and offers opportunities to transfer knowledge to new problems or situations (McDaniel, Roediger, & MacDermott, 2007; Butler & Roediger, 2007). Second, formative assessment may be effective because feedback further enhances knowledge retention. A number of studies have demonstrated that well-designed cycles of testing with feedback and opportunities for knowledge updating can support students in practice that leads to mastery of the desired skills and concepts (e.g., Pavlik, et al., 2007).

The present investigation focuses on evaluating the impact of spaced practice and formative assessment for individual learners as applied to specific aspects of mathematics knowledge and skills. We are examining the extent to which an online tutor designed to automatically reassess student knowledge and provide relearning opportunities improves long term retention of knowledge for individual students and across specific mathematical skills.

Methods

Using an online platform called ASSISTments (Feng, Heffernan, Koedinger, 2009), we conducted an empirical study to investigate spaced practice and formative assessment. ASSISTments is a domain-general web-based system that allows teachers to create individual practice and assessment assignments, composed of questions and associated hints, solutions, web-based videos, and the like. The participants included 97 8th grade students who completed the required pre- and post-tests in the classrooms of the two cooperating teachers. Both teachers were experienced using the CMP2 curriculum and using the ASSISTments platform. The research setting was a suburban middle school in Massachusetts in a relatively high SES district.

A total of 32 skills were targeted for study and were divided into two sets, Set A and Set B (see Table 1). The skills varied with regards to when they should have previously been learned. Each student was provided opportunities for practice and assessment on each skill, and the separation into two skill sets was for purposes of counterbalancing the assignment of skills to the two testing-practice treatments across student groups (see Figure 1).

A 3-factor mixed design was used. The within-subjects factors were the Practice Condition (Mastery-learning only, ARRS) and Test (Pretest, Posttest). The between-subjects factor was Group (1, 2), which corresponded to which skill set students received in each practice condition (see Table 2). This counterbalanced design allows every student to engage in both Practice Conditions.

Data Sources

In September 2010, students were assigned five skills per week from the total pool of skills, to be completed online as part of their homework. Students received one pretest item for each of the 32 skills. The pretest item for each skill also served as the first item in the mastery learning set. In mastery learning, students were required to answer three problems correct in a row for each skill. Therefore, if a student correctly answered the pretest item and the next two problems, that student was determined to have shown "mastery" of that skill. If an item was answered incorrectly, the student was provided feedback (i.e. correctness and hints) to help solve the problem. The student would be given additional practice until he or she was able to answer three consecutive problems correct for the skill without the use of feedback.

After a student demonstrated mastery on a particular skill, that skill entered one of two conditions (see Figure 1). Either the skill was placed in the control condition or the ARRS condition. In the "mastery-only" control condition, the skill received no further assessment and practice. The ARRS condition consisted of two parts. First, students were reassessed on their mastery of the skill. That is, students were given reassessment opportunities where they had to demonstrate retained mastery of the skill by answering one problem correctly. Skills were reassessed on a schedule with increased spacing intervals of 7, 14, 30, and 60 days. If the student demonstrated retention on Day 7, the skill was reassessed 14 days later; if the student demonstrated retention at the 14-day reassessment interval, the skill was reassessed after 30 days.

Relearning is the second component of the ARRS condition. When a student did not demonstrate retention, feedback was given and practice continued until the student could demonstrate mastery by answering three correct questions in a row for a skill. Additionally, the reassessment interval only increased if the student demonstrated retention at that reassessment interval. For example, if a student could not demonstrate retention at the 7-day interval, practice with feedback would be provided until the student re-mastered the skill. However, the student would remain at the 7-day interval until relearning was no longer required for that skill. In both conditions, a posttest measuring retention performance for all skills was administered in March 2011.

Results

This study investigated the degree of benefit of ARRS relative to the "baseline" mastery condition, There was a Condition x Test interaction (see Figure 2), $F(1, 95) = 13.50$, $p < .001$. There was no difference between the ARRS (M =.60, SE = .02) and the Mastery-only (M =.61, SE = .02) condition in the pretest, $F < 1$, *ns*. However, post-test performance in the ARRS condition (M =.74, SE = .02) was significantly better than Mastery-only (M =.67, SE = .02), $F(1, 95) = 13.79$, $p < .001$. Figure 3 is a bivariate data plot of pretest and posttest scores for the 97 students with each student contributing one score for ARRS items and one for Mastery-only items (194 data points). We conducted a regression analysis using condition, pretest score, and the interaction term (pretest x condition) as predictors of posttest performance for individual students. The full model containing the three variables explained 39.1% of the variance ($F(3,190) = 40.582$, $p<.001$). Controlling for pretest scores, students' ARRS condition scores (M = .77) were significantly higher at posttest than their Mastery-only posttest scores, (M = .69), t (190) = 4.04, $p < .001$. Pretest significantly predicted posttest performance, t (190) = 6.99, $p < .001$. There was no significant interaction between pretest and condition, t (190) = 1.34, p = .18.

We were also interested in comparing the relative value of ARRS for knowledge and skills that are more versus less challenging. Figure 4 shows the bivariate distribution of performance on individual items averaged across Ss based on item treatment assignment. A regression analysis indicated that the overall model with 3 factors explained 56.4% of the variance (F(3,60)

 4

= 25.924, p <.001). Moreover, controlling for pretest scores, problems assigned to the ARRS condition (*M* = .788) were significantly higher at posttest than when assigned to the Mastery-only condition, (*M* = .712), t (60) = 2.564, p < .01. Pretest scores significantly predicted posttest performance for individual problems, t (60) = 5.52, p < .001. There was no significant interaction between pretest and condition, t (60) = .63, p = .53.

Scholarly significance of the work

This study, which applies the instructional design principles of spaced practice and formative assessment, has important implications regarding the use of technology in the creation of effective learning environments. Our results show that ARRS is an effective method for increasing student retention of critical knowledge and skills and in maintaining high levels of proficiency. When skills are reassessed and relearned, students perform better on those skills relative to skills that were practiced and mastered but not further reassessed and relearned. Not surprisingly, ARRS was especially beneficial for students who performed lower on the pre-test and for more difficult and less well retained skills but it also showed benefits across the entire student and skill performance range.

There are numerous advantages of using a system like ASSISTments to automatically practice, reassess and provide relearning opportunities. First, providing students with individualized spaced practice and testing with feedback is a difficult and complex task without the use of technology. Providing students with such practice and feedback would not be possible with traditional paper and pencil curricular materials. Second, automating the process of assigning practice of specific skills based on individual student performance allows for more meaningful and tailored practice and relearning. Students spend more time practicing difficult skills and spend less time practicing those they have already mastered. Finally, teachers can use feedback generated by ASSISTments to determine skills and concepts students find difficult and tailor classroom instruction appropriately. Further educational and theoretical implications will be discussed.

References

Bahrick, H. P., & Hall, L. K. (1991). Lifetime maintenance of high school mathematics content. *Journal of Experimental Psychology: General, 120*, 20–33.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74.

Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*, 514-527.

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.

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply results of psychological research. *American Psychologist, 43*, 627–634.

Feng, M., Heffernan, N.T., & Koedinger, K.R. (2009). Addressing the assessment challenge in an intelligent tutoring system that tutors as it assesses. *User Modeling and User-Adapted Interaction. 19,* 243-266.

Karpicke, J. D., & Roediger, H. L., III. (2010). Is expanding retrieval a superior method for learning text materials? *Memory & Cognition, 38,* 116-124.

Klahr, D., & Nigram, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction on discovery learning. *Psychological Science, 41*(2). 75-86.

Lappan, G., Fey, J. T., Fitzgerald, W. M., Friel, S. N., & Phillips, E. D. (2009). *Connected Mathematics 2*. Pearson: Boston.

Mayfield, K. H., & Chase, P. N. (2002). The effects of cumulative practice on mathematics problem solving. *Journal of Applied Behavior Analysis, 35,* 105–123.

McDaniel, M., Roediger, H. L., & McDermott, K. (2006). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin and Review, 14*, 200-206.

Nyquist, J. B. (2003). The benefits of reconstructing feedback as a larger system of formative assessment: A meta-analysis. Unpublished master's thesis, Vanderbilt University, Nashville, TN.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., et al. (2007). *Organizing instruction and study to improve student learning. IES practice guide* (NCER 2007-2004). Washington, DC: National Center for Education Research. Retrieved February 28, 2008, from http://ies.ed.gov/ncee/wwc/pdf/practiceguides/20072004.pdf

Pavlik, P. I., Jr., Presson, N., Dozzi, G., Wu, S.-M., MacWhinney, B., & Koedinger, K. (2007). The FaCT (fact and concept) system: A new tool linking cognitive science with educators. In *Proceedings of the 29th Annual Conference of the Cognitive Science Society*. Nashville, TN, USA.

Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Current Directions in Psychological Science, 16*, 183-186.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics practice problems improves learning. *Instructional Science*, *35*, 481-498.

Rohrer, D. (2009). The effects of spacing and mixing practice problems. *Journal for Research in Mathematics Education*, *40*, 4-17.

Underwood, B. J. (1961). Ten years of massed practice on distributed practice. *Psychological Review, 68*(4), 229-247.

Willingham, D. T. (2002, Summer). Allocating student study time: Massed vs. distributed practice. *American Educator,* 37-39, 47.

Table 1. Skill sets A and B

Skill Set A	Skill Set B
Compute the corresponding lengths between scale objects.	Determine if two algebraic expressions are equivalent.
Divide fractions.	Multiply and divide integers.
Convert decimals into fractions.	Determine divisibility by a number.
Multiply fractions and mixed numbers.	Solve a word problem by determining the least common multiple.
Given a diagram, compute the perimeter of a polygon.	Write algebraic expressions to model situations.
Evaluate numeric expressions with absolute value.	Compute the percent of a number.
Determine the absolute value of a number.	Add positive and negative decimals.
Evaluate numeric expressions by applying the order of operations.	Convert between scientific notation and standard form.
Convert fractions into percents.	Determine the least common multiple of two numbers.
Add and subtract fractions and mixed numbers.	Combine two algebraic expressions by adding.
Solve for an unknown in a proportion.	Combine two algebraic expression using substitution.
Use substitution to simplify algebraic expressions.	Given an equation, determine whether it represents the distributive, associative, or commutative property.
Given the radius of a circle, compute the area.	Find the prime factorization of a number.
Solve algebraic equations for a variable in terms of another variable.	Determine the greatest common factor of two numbers.
Given a diagram, compute the area of an irregular figure.	Simplify expressions using the distributive property.
Compute discount and sales tax.	
Solve algebraic equations with one variable.	

 8

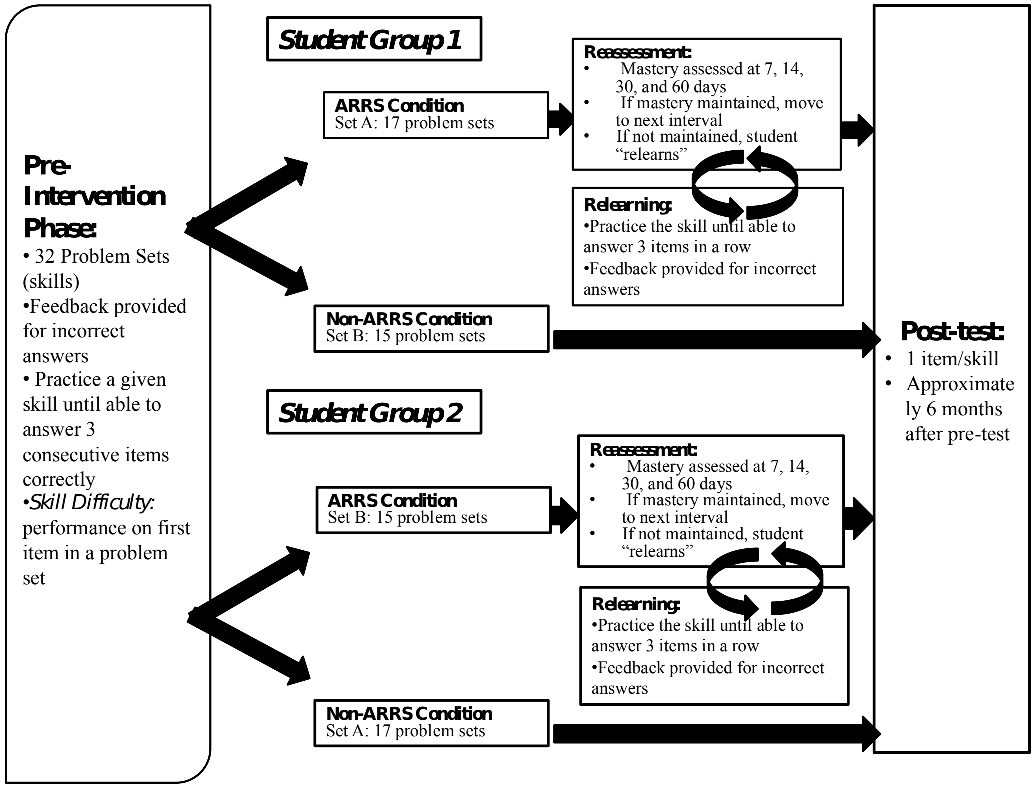

Figure 1. Design of the study

Table 2. Counterbalancing of Skill Sets across Practice Conditions

	ARRS	Mastery-only
Group 1	Skill Set A	Skill Set B
Group 2	Skill Set B	Skill Set A

Figure 2. The effect of practice condition (ARRS vs. Mastery-learning only) as a function of pretest and posttest performance.

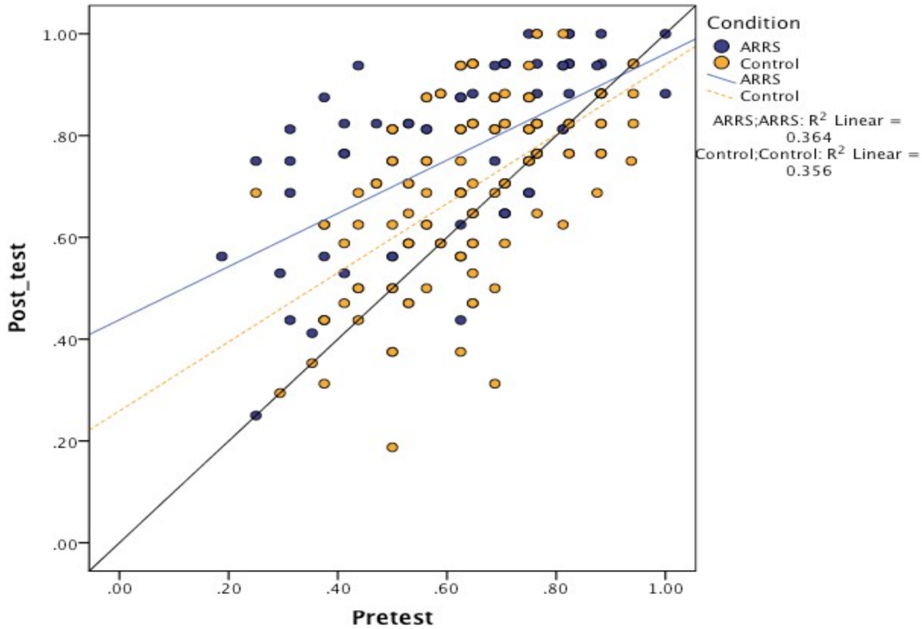

Figure 3. The bivariate distribution of individual student pretest and posttest scores as a function of practice condition. Least squares regression lines are plotted for each practice condition.

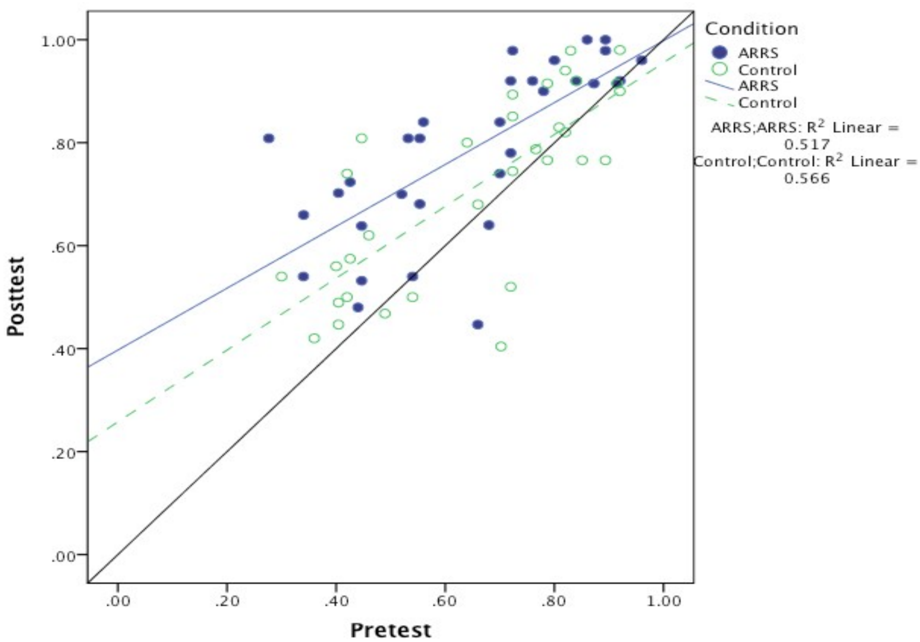

Figure 4. The bivariate distribution of individual mathematics item pretest and posttest scores as a function of practice condition. Least squares regression lines are plotted for each practice condition.