

Abstract Title Page.

Title: Cognitively-Based Instructional Design Principles: A Technology for Testing their Applicability via Within-Classroom Randomized Experiments

Author(s): Neil Heffernan, Cristina Heffernan, Kevin Dietz, Deena Soffer, James W. Pellegrino, Susan R. Goldman, & Matthew Dailey

Abstract Body

Background / Context:

The work we propose to report is part of the program of research and development underway at the IES-funded *National Center on Cognition and Mathematics Instruction*. The Center team is applying research-based design principles to revise mathematics curricular materials for the critical grade span of 6-8, when fundamental concepts required for success in algebra and later math courses are addressed. In the course of the redesign, the Center team is testing hypotheses that the design principles can individually and collectively contribute to substantial improvements in student learning and engagement with critical topics in mathematics.

The *IES Practice Guide* (Pashler et al., 2007) provides a number of research-based recommendations for improving instruction. These recommendations are stated in a sufficiently “generic” form to be potentially applicable to a wide range of instructional content areas and curricular materials. The Center is (1) translating those principles into specific practices applicable to the materials of the Connected Mathematics Project (CMP) curriculum designed for grades 6-8; and (2) testing the efficacy of the principle-based revisions.

One of the instructional practices recommended in the *IES Practice Guide* is the spacing of practice and learning over time (Rec. 1). Extensive research has demonstrated large retention advantages when learners are exposed to key facts, concepts, and knowledge at multiple points in time (see Cepeda, Pashler, Vul, et al, 2006), a phenomenon dubbed the “spacing effect.” When learners practice recalling and applying relevant information, they are more likely to retain that knowledge for a greater period of time (Rohrer & Taylor, 2007; Rohrer & Pashler, 2007). In a review of the spacing effect literature, Rohrer (2009) argued, “the spacing of practice is being grossly underutilized in mathematics instruction”, as other authors have also concluded (e.g., Bahrick & Hall, 1991; Dempster, 1988; Mayfield & Chase, 2002; Willingham, 2002).

The spacing effect is complementary to a second instructional design principle emphasized in the *IES Practice Guide*, focused on quizzing and assessment (Rec. 5). Research on the so-called “testing effect” (e.g., Butler & Roediger, 2007; McDaniel, Roediger, & MacDermott, 2007) suggests that such activities are successful compared to restudy in part because they prompt students to retrieve information, reflect on the state of their knowledge based on feedback, and they offer opportunities to transfer knowledge to new problems or situations. Well-designed cycles of testing with feedback with opportunities for knowledge updating can support students in mastery and retention of desired skills (e.g., see the FaCT system of Pavlik et al., 2007).

The investigation we will present is focused on evaluating the impact of variations of the “spacing of practice” and “testing effects” principles as applied to specific aspects of mathematics knowledge and skill. Our emphasis was on skills that are supposed to be mastered (and retained) over the course of a single academic year, as well as those skills ostensibly mastered in prior years and that serve as prerequisite knowledge to success in new learning units.

Purpose / Objective / Research Question / Focus of Study:

Redesign for optimal spacing of practice, as well as assessment with performance feedback, requires that there be a clear and consistent definition of the “what” that is being practiced and/or assessed and when it is introduced, practiced, assessed and ostensibly mastered over the three year span of the CMP curriculum. Our first step in implementing principles of spacing and testing was to identify the “what” and then trace its occurrences from initial mastery to subsequent opportunities for practice and assessment. The “what” are knowledge components (skills, context, procedures, and responses). We have identified a total of 148 skills and about 70

prerequisite relationships among them. We are mapping these within and across units from grade 6 to grade 8 by tagging all practice and test questions in the CMP curriculum.

Understanding our approach to tracing these skills over the curriculum requires an explanation of the structure of units in the curriculum itself. Each Unit in CMP is divided into 3 to 5 Investigations and each Investigation is comprised of 2 to 5 lessons. Following the completion of each lesson, the *Teacher's Guide* provides recommended ACE (Applications-Connections-Extensions) homework problems to assign to students. Homework problems are identified by whether they practice concepts learned in the most recent lesson (Applications), review previously learned material (Connections), or provide challenging problems beyond the concepts already taught in the curriculum (Extensions).

Through our analysis of the CMP curricular materials, we are able to ascertain the practice frequency of specific skills as well as relationships between the practice problems and the assessment problems included in the CMP materials. Through use of a platform known as ASSISTments (Feng, Heffernan, Koedinger, 2009), we are conducting empirical studies to analyze optimal spacing and testing as they relate to the difficulty of a skill and how different types of feedback affect its retention. 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 etc. It was developed at WPI and further information about its components and use can be found at www.ASSISTments.org.

The specific study we conducted was focused on comparing the relative and absolute benefits of two different practice-testing conditions on the retention of skills and knowledge that are essential targets of the CMP middle-grades math curriculum. In all cases, students had received instruction on the skills. Then, in both conditions, in the context of homework activities, students had to first demonstrate mastery by getting three items correct in a row on a given skill. For those skills where students made mistakes they were given assistance in the form of hint messages. Once mastered, for each student half of the skills underwent distributed retesting with additional practice if needed. All skills were then evaluated subsequently for levels of long-term retention.

Setting / Population / Participants / Subjects:

The research setting was a suburban middle school in Massachusetts that is part of a relatively high SES district with experience using the ASSISTments platform. Two teachers in the school assisted in conducting the research with the active support of their principal who was the first teacher to ever use ASSISTments when it was created eight years ago. The participants included 128 students who completed the required pre- and posttests in the 8th grade classrooms of the two cooperating teachers, both of whom were experienced with the CMP curriculum which was being used by all their students.

Intervention / Program / Practice:

A total of 33 skills were targeted for study and they were divided into two sets (A and B) as shown in Table 1. The skills varied in when they should have previously been learned as well as the extent to which they were prerequisite for acquiring new content in a CMP unit that was the focus of instruction shortly after the study began. Each student was provided opportunity for testing and practice on each skill and the separation into two skills sets was for purposes of counterbalancing the assignment of skills to the two testing-practice treatments across student groups as shown in Figure 1. Each student received each treatment.

Starting in September 2010, students were assigned five skills per week from the total pool of skills, to be completed online for homework. The ASSISTments web-based system was used to present the students with mathematics problems that tested the specific skills. Over four months, the students revisited each skill on a testing and practice schedule designed to determine whether they had retained knowledge of the previously “learned” skill. Students were determined to have shown “mastery” on a skill when they correctly completed three consecutive problems tapping that particular skill. If the student answered incorrectly, he/she was required to practice that skill again until attaining mastery. For purposes of analysis, performance on the initial problem assessing a given skill was used as a pretest measure and performance over the first three practice problems was used as a measure of baseline performance (see below).

Each student participated in two conditions of practice and testing: (a) The Automatic Reassessment and Relearning System (ARRS), which required that during “reassessment” the students demonstrate “mastery learning” (i.e. three consecutive correct answers), which was then followed up with distributed reassessment and relearning opportunities and (b) Non- Automatic Reassessment and Relearning System, in which students were also required to demonstrate “mastery learning” but in which they were not provided with subsequent reassessment and relearning opportunities. Thus, once mastered, the student did not see the skill again until the final posttest. In contrast, in the ARRS condition, skills were reassessed on a schedule such that students were required to demonstrate proficiency at increased spacing intervals of 7 days later, 14 days after that, 30 days after that, and then 60 days after that. If the student answered correctly on day 7, the skill was reassessed again 14 days later; if the student answered correctly at the 14 day reassessment interval, the skill was reassessed at 30 days; likewise at 60 days. However, if the student answered incorrectly, the student was required to re-master the skill and the skill was reassessed 7 days later (i.e., the skill kept the same initial reassessment interval). In other words, the reassessment interval only increased if the student demonstrated mastery retention at that reassessment interval. In all cases, when a skill did not demonstrate retention, feedback was given and practice continued until three correct in a row. A final posttest measuring retention performance for all skills was administered in March 2011.

Research Design:

A 3-factor mixed between-Ss and within-Ss design was used. The between-subjects factor was Group (Group 1, Group 2). The within-subjects factors were the Practice Condition (ARRS, Non-ARRS) and Test (Pre-test vs Post-test; or Baseline vs Post-test). As shown in Figure 1, the Group factor counterbalances the assignment of sets of math skills to the two different Practice Conditions. Thus, the Practice Condition variable was tested within-Ss and any interaction between Group and Practice Condition constitutes a test of the main effect of Skill Set.

Data Collection and Analysis:

The current study investigated the extent to which ARRS is more beneficial to skill retention than a condition where mastery learning is obtained but not reassessed (Non-ARRS). The data were analyzed in two separate ways. First, the data were analyzed using a 2x2x2 mixed Analysis of Variance (ANOVA) design in which the Test variable compared how student performance on ARRS and Non-ARRS practiced skills changed from pre-test to post-test. A second 2x2x2 mixed Analysis of Variance (ANOVA) design examined how the ARRS and Non-ARRS practiced skills were retained on the posttest relative to initial baseline performance.

Findings / Results:

The first analysis examining student performance at both pretest and posttest revealed several main effects and interpretable interactions. As might be hoped, students answered more questions correctly on the post-test ($M = .68$ $SD = .20$) than on the pre-test ($M = .46$ $SD = .17$), $F(1, 126) = 228.12, p < .001, \eta_p^2 = .64$ (.8 effect size). In addition, skills originally mastered and then practiced under the ARRS condition were answered more correctly ($M = .59$ $SD = .18$) than skills originally mastered under the non-ARRS condition ($M = .55$ $SD = .17$), $F(1, 126) = 19.58, p < .001, \eta_p^2 = .13$ (.36 effect size). Both main effects can be seen in Figure 2 which also shows the Practice Condition x Test interaction, $F(1,126) = 11.58, p < .01, \eta_p^2 = .08$ (.28 effect size). Follow-up tests of the Practice Condition x Test interaction revealed that ARRS and non-ARRS skills did not differ in performance at pre-test, $F < 1, ns$. In contrast, students performed significantly better at posttest on ARRS practiced skills than non-ARRS skills, $F(1, 127) = 29.22, p < .001, \eta_p^2 = .19$ (.44 effect size) (See Figure 2).

A Practice Condition x Group interaction was also observed, $F(1,126) = 45.54, p < .001, \eta_p^2 = .27$ (.52 effect size). As noted earlier, this interaction is actually a main effect of skill set. As shown in Figure 3, skill set B was more difficult overall than skill set A. There was no main effect of Group, $F < 1, ns$, and no Group x Practice Condition x Test interaction, $F < 1, ns$.

The second analysis examined post-test retention performance relative to baseline performance levels obtained during initial testing and mastery practice. This analysis also revealed several main effects and interpretable interactions. Overall, performance was better on skills practiced under the ARRS condition ($M = .71$ $SD = .16$) than skills in the non-ARRS practice condition ($M = .67$ $SD = .16$), $F(1, 126) = 24.72, p < .001, \eta_p^2 = .16$ (.40 effect size). As shown in Figure 4, there was a Practice Condition x Test interaction, $F(1,126) = 19.02, p < .001, \eta_p^2 = .13$ (.36 effect size). Follow-up tests revealed that ARRS practiced skills maintained their level of accuracy from baseline mastery to posttest, $F < 1, ns$. In contrast, non-ARRS practiced skills decreased from baseline mastery to posttest, $F(1,127) = 14.95, p < .001, \eta_p^2 = .11$ (.33 effect size) (see Figure 4). Thus, the main effect of Practice Condition is solely a function of differences between ARRS and non-ARRS at Posttest.

A Practice Condition x Group interaction was again observed, $F(1,126) = 30.69, p < .001, \eta_p^2 = .20$ (.45 effect size). As shown in Figure 5, this is actually a main effect of skill set indicating that skill set B is more challenging than skill set A. There was no main effect of Group, $F < 1, ns$, and no Group x Practice Condition x Test interaction, $F(1,126) = 2.19, ns$.

Conclusions:

Based on our results, ARRS appears to be an effective method for increasing student performance over time and maintaining high levels of mastery. Our results clearly show that when skills are reassessed and relearned, students perform nearly half a letter grade (approximately 5 percentage points) better than students who were not reassessed and did not have relearning opportunities. Additionally, there is evidence to suggest that ARRS is beneficial to maintaining performance levels. Students using ARRS were able to maintain their performance levels from baseline mastery to post-test, whereas, students without the benefit of ARRS distributed practice and assessment showed a decrease in their performance from baseline mastery to posttest. The study also shows that many of the skills presumed to have been mastered in earlier years and lessons are in a very weak state of proficiency (pretest average of .46) and need periodic reassessment and practice. We hope to be able to test optimal schedules for doing so by using ASSISTments which permits random assignment within classrooms.

Appendices

Appendix A. References

- Bahrick, H. P., & Hall, L. K. (1991). Lifetime maintenance of high school mathematics content. *Journal of Experimental Psychology: General*, *120*, 20–33.
- 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 the 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.
- 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.
- 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.
- Willingham, D. T. (2002, Summer). Allocating student study time: Massed vs. distributed practice. *American Educator*, *37–39*, 47.

Appendix B. Tables and Figures

Table 1. Specific skills assigned to Skill Sets A and B

| Skills A | Skills B |
|---|---|
| <ul style="list-style-type: none"> • Absolute Value • Absolute Value, Addition, Subtraction • Addition and subtracting Fractions • Area of Circle Pre • Area of Irregular Figure Pre • Conversion - Fraction to Percent • Converting fraction to percent • Converting fractions to decimals • Discounts and Sales Tax • Dividing Fractions • Greatest common Factor • Multiplication Fractions • Order of operations - basic Pre • Perimeter • Scale Drawing • Solving equations Pre • Solving for unknown using scale factor • Solving variable equation Pre • Substitution Pre | <ul style="list-style-type: none"> • Addition Decimals • Composition of functions adding Main • Composition of functions substitution Main • Define Distributive Commutative Associative Main • Distributive Property Pre • Divisibility • Equations from a Diagram • Least Common Multiple • Least Common Multiple in a Word Problem • Multiplication and Division Integers • Percent of • Prime Factorization • Recognizing Linear Equivalent Expressions Main • Scientific Notation |

| | | |
|---------|-------------|-------------|
| | ARRS | Non-ARRS |
| Group 1 | Skill Set A | Skill Set B |
| Group 2 | Skill Set B | Skill Set A |

Figure 1. Counterbalancing of Skill Sets across Practice Conditions.

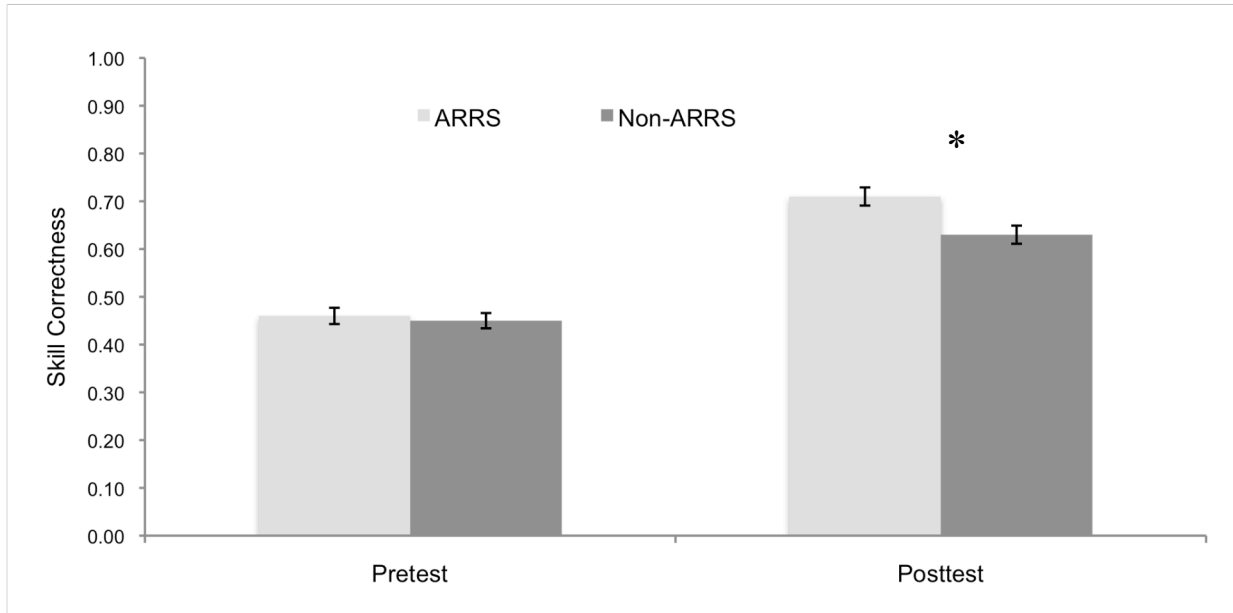


Figure 2. Mean accuracy on math skill items by Test and Practice Condition. Error bars represent standard error. $p < .001$.

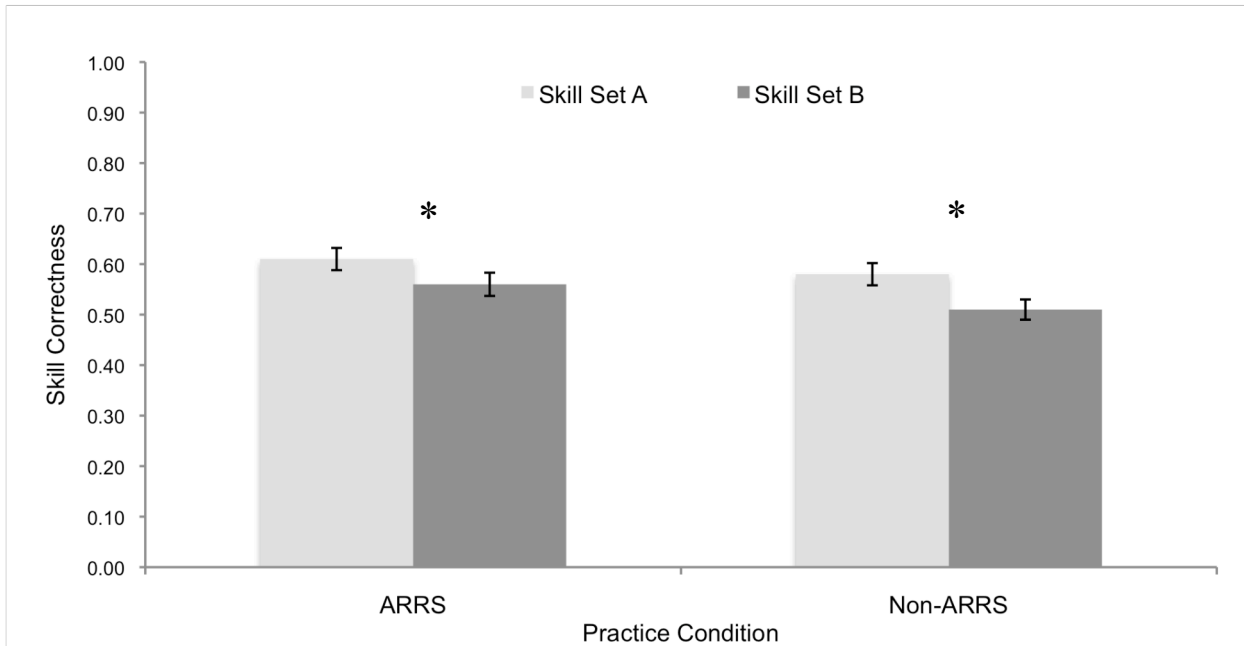


Figure 3. Mean accuracy on math skill items by Practice Condition and Skill Set collapsing across Test (Pretest, Posttest). Error bars represent standard error. $p < .05$

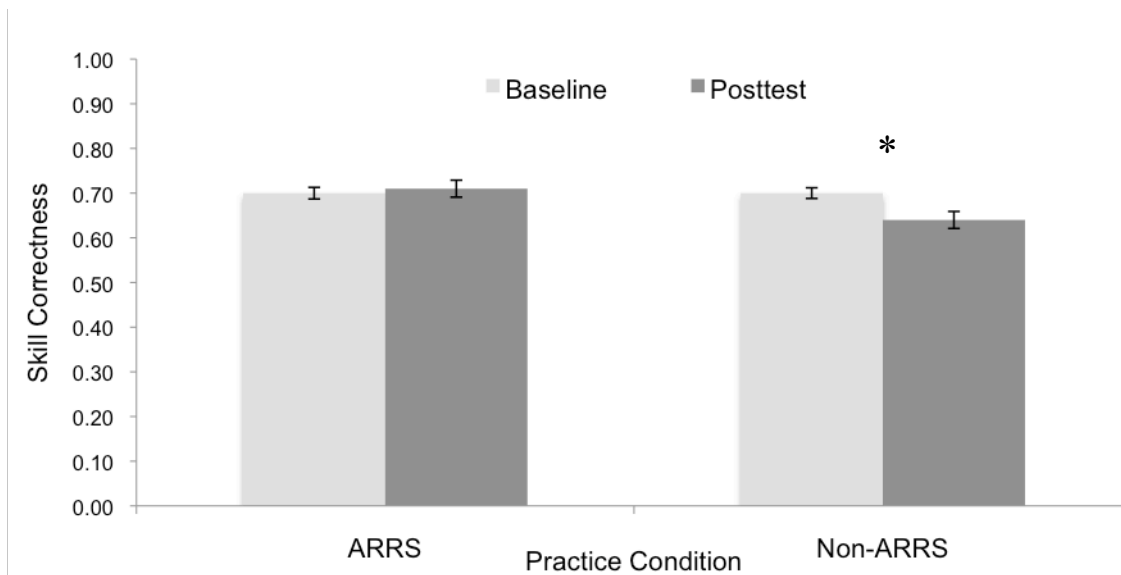


Figure 4. Mean accuracy on math skill items by Practice Condition and Mastery Baseline vs Posttest. Error bars represent standard error. $p < .001$.

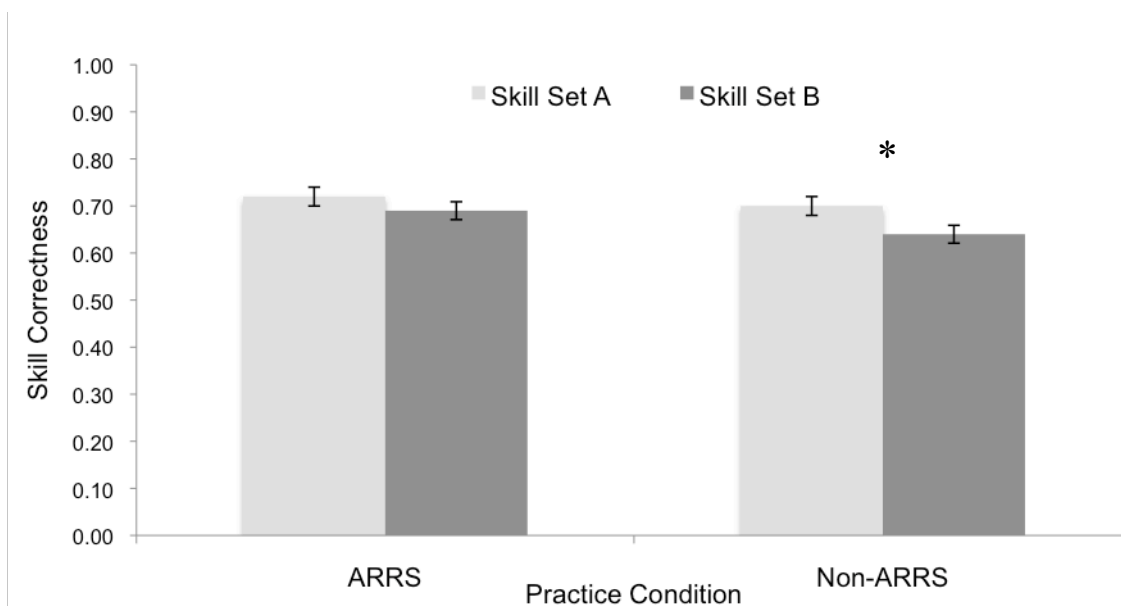


Figure 5. Mean accuracy on math skill items by Practice Condition and Skill Set when collapsing across Test (Baseline Mastery, Post-test). Error bars represent standard error. $p < .05$