Journal of Education for Students Placed at Risk (JESPAR)

ISSN: 1082-4669 (Print) 1532-7671 (Online) Journal homepage: http://www.tandfonline.com/loi/hjsp20

Design-Based Research Within the Constraints of Practice: AlgebraByExample

Julie L. Booth, Laura A. Cooper, M. Suzanne Donovan, Alexandra Huyghe, Kenneth R. Koedinger & E. Juliana Paré-Blagoev

To cite this article: Julie L. Booth, Laura A. Cooper, M. Suzanne Donovan, Alexandra Huyghe, Kenneth R. Koedinger & E. Juliana Paré-Blagoev (2015) Design-Based Research Within the Constraints of Practice: AlgebraByExample, Journal of Education for Students Placed at Risk (JESPAR), 20:1-2, 79-100, DOI: 10.1080/10824669.2014.986674

To link to this article: http://dx.doi.org/10.1080/10824669.2014.986674

Published online: 31 Mar 2015.

Journal of Education for Students Placed At Risk, 20: 79–100, 2015

ISSN: 1082-4669 print / 1532-7671 online
DOI: 10.1080/10824669.2014.986674

Design-Based Research Within the Constraints of Practice: AlgebraByExample

Julie L. Booth
Department of Psychological, Organizational, and Leadership Studies, Temple University

Laura A. Cooper
Evanston Township High School District 202

M. Suzanne Donovan and Alexandra Huyghe
Strategic Education Research Partnership

Kenneth R. Koedinger
Human-Computer Interaction Institute, Carnegie Mellon University

E. Juliana Paré-Blagoev
Strategic Education Research Partnership

Superintendents from districts in the Minority Student Achievement Network (MSAN) challenged the Strategic Education Research Partnership (SERP) to identify an approach to narrowing the minority student achievement gap in Algebra 1 without isolating minority students for intervention. SERP partnered with 8 MSAN districts and researchers from 3 universities to design and rigorously test AlgebraByExample, a set of 42 Algebra 1 assignments with interleaved worked examples that target common misconceptions and errors. In a year-long random-assignment study, students who received AlgebraByExample assignments had an average 7 percentage point boost on a posttest containing released items from state assessments, and students in the bottom half of the performance distribution where minority students are disproportionately concentrated had an average 10 percentage point boost on a researcher-designed assessment of conceptual understanding. AlgebraByExample is easily incorporated into any existing curriculum, and naturally serves as a launch point for mathematically rich discussion.

The Minority Student Achievement Network (MSAN) is a growing network of 29 suburban and small urban districts committed to eliminating the achievement gaps between their White and Asian students and their African American and Latino students (who range in number from 20% to 83% of the school population). The network was self-organized to accelerate learning in literacy and math and improve college-going rates by learning from each other and from

---

Correspondence should be addressed to M. Suzanne Donovan, 1100 Connecticut Avenue NW, Suite 1310, Washington, DC 20036. E-mail: sdonovan@serpinstitute.org

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hjsp.

research. The course Algebra 1 was of particular concern to these districts because it plays an important gatekeeper role with regard to higher-level, college-preparatory mathematics and because the districts were not making progress in accelerating the learning of their African American and Latino students, despite the districts' focus on improving curriculum and instruction to incorporate the current best thinking. The network leaders looked to researchers for new ideas.

MSAN recognized that for practitioners with full workloads, the demands of forming and managing a productive research–practice partnership are unrealistic. Bridging the worlds of research and practice requires identifying and recruiting researchers who have both the expertise and interest in collaborative, problem-solving research and development, then launching and managing a productive program of work. When the Strategic Education Research Partnership (SERP) was formed to play precisely this role, MSAN leaders saw an opportunity.

SERP, an independent, nonprofit organization incubated at the National Academy of Sciences, was established with a mission to conduct education research and development in Pasteur's quadrant (Stokes, 1997), where new knowledge is generated in the interest of solving important problems of practice. SERP work is carried out at field sites where partnerships between school district practitioners and interdisciplinary research and design teams are created, nurtured, and, funding permitting, sustained. Participating districts define the focal problem, and SERP-recruited researchers help to frame the problem with respect to knowledge from prior research. Researchers and practitioners collaborate at every stage so that the work is not only research-informed, but also reality-checked. MSAN and SERP formed a partnership in 2006 to address the Algebra 1 challenge.

FOCUS ON ALGEBRA 1

Although SERP work responds to the needs of district partners, expectations are established at the outset regarding the ultimate goal of producing knowledge and tools that are of benefit not only to the partner district, but to the field of education more broadly. The Algebra 1 achievement gap certainly meets that criterion. Substantial discrepancies in the mathematics achievement levels of students by race and ethnicity persist, despite increased attention to the issue. Results from the most recently available National Assessment of Educational Progress revealed that the Black–White mathematics achievement gap for eighth graders, a grade in which Algebra 1 is often taught, was 31 points in 2007 (Vanneman, Hamilton, Anderson, & Rahman, 2009). In 2009, the achievement gap between Latino and White students was 26 points, and was even greater when comparing White students to Latino English language learners (Hemphill & Vanneman, 2011).

Algebra 1 can be particularly challenging not only because it introduces more abstract representations and more complex mathematical relationships, but also because it can magnify misconceptions that have their roots in earlier instruction. Further, given the age at which students first take an Algebra 1 course, leaders in the MSAN districts feared that emerging self-consciousness could lead students who perform poorly to perceive themselves as incompetent in the domain, and become less mastery-oriented and more avoidance-motivated (Harter & Connell, 1984). Although the drive to avoid failure is not uncommon in any domain, students who select themselves out of mathematics courses at the earliest possible opportunity relinquish access to a broad range of higher education and career opportunities.

THE APPROACH

SERP recruited an interdisciplinary team of researchers to work with practitioners from an initial subset of five MSAN districts to frame the problem more precisely and define an approach to addressing it. Mathematics researchers and program designers were selected to bring a diverse set of ideas to bear on potential approaches to addressing the problem; among them were contributors to the development of four different, innovative mathematics curricula.

In accordance with SERP protocols, superintendents and instructional leadership in the participating districts defined the parameters for the work. On the basis of decades of experience, the district leaders set three constraints:

1. A supplemental approach introduced in afterschool hours or during the summer was undesirable because, in the superintendents' experience, such programs are the first to disappear during budget cuts.
2. An entirely new curriculum was not an option because both the political and financial costs of such a decision would be high, and teachers' focus would be directed to the demands of the new curriculum rather than to struggling students.
3. The problem could not be addressed by explicitly or uniquely targeting minority students or struggling students because this would reinforce any tendency for these students to self-identify as being "not good at math." Empirical work has demonstrated that such targeting can increase minority students' stereotype threat, or "risk of confirming a negative stereotype about one's group," with inimical consequences for performance attainment (Steele & Aronson, 1995, p. 797). The districts wanted an approach that would target all students, even if it was intended to be particularly beneficial for a subset of the population.

SERP balances the interests of practitioners with the interests of researchers; approaches to a problem must be based in learning principles for which there is research evidence, and be tested using rigorous methodologies. And SERP requires that any solution to a problem be scalable. If scalability requires that teachers internalize a change and take ownership (Coburn, 2003), then it is particularly important to design and test to ensure that teachers see the approach as both worthy and workable in their classrooms.

The SERP-recruited researchers and district practitioners thought through approaches to improvement that fit within these multiple constraints. As is often the case, those who are further down in the district hierarchy brought insights about obstacles to improvement that were encountered closer to the ground and were not seen by the most senior district leaders. District math coordinators emphasized that top-down directives requiring teachers to substantially change the way they teach were not going to work because algebra teachers in their districts believe they know more about mathematics teaching and about their students than anyone in the district's central office. Furthermore, they argued that teachers would resist making changes that diminish their sense of control, a phenomenon documented in field research by Mary Kennedy (2005).

A math coordinator in one of the districts recounted a recent experience in which a practice in which teachers showed no interest was introduced in summer school where the stakes are lower, and later spread into the school year. He argued that the likelihood of success in improving Algebra 1 learning would be highest if a *back door* was identified—an unobtrusive approach that would demonstrate the benefit of a new practice without having the primary routines of current practice upended. Ken Koedinger, a SERP-recruited researcher from Carnegie Mellon University, proposed an approach that he and his colleagues found to be powerful in experimental research: providing students with worked examples and self-explanation prompts (cf. Pashler et al., 2007). Worked examples were especially appealing because they could be easily introduced through assignments. This approach would not require significantly disrupting a teacher's instructional routines, but has the potential to increase the efficacy of teacher and student work. The approach was not intended to be teacher-proof. In fact, the hope was that teachers would review assignments as a classroom activity, and thus be exposed to students' thinking and to the value of reasoning through worked examples.

Over the next 7 years, SERP staff members and researchers collaborated with practitioners across eight MSAN districts to iteratively develop and test AlgebraByExample, a set of 42 math assignments that incorporate powerful research findings regarding student misconceptions and the value of self-explanation.

ALGEBRABYEXAMPLE RESEARCH BASE

Many students enter the Algebra 1 classroom holding misconceptions that have the strong potential to derail new learning (Brown, 1992; Chiu & Liu, 2004; Kendeou & van den Broek, 2005). Within the domain of equation-solving alone, a number of misconceptions have been identified as critical, including the idea that the equals sign is an indicator of operations to be performed (Baroody & Ginsburg, 1983; Kieran, 1981; Knuth, Stephens, McNeil, & Alibali, 2006); that negative signs represent only the subtraction operation and do not modify terms (Vlassis, 2004); that subtraction is commutative (Warren, 2003); and that variables cannot take on multiple values (L. R. Booth, 1984; Knuth et al., 2006; Küchemann, 1978). Not surprisingly, such misconceptions have been shown to affect students' success in problem solving and hinder their learning of new material (J. L. Booth & Koedinger, 2008). For many students, these misconceptions persist even after traditional classroom instruction on the relevant topic (J. L. Booth, Koedinger, & Siegler, 2007; Vlassis, 2004). A persuasive body of evidence from a variety of content areas suggests that dislodging misconceptions can be devilishly difficult, and often requires directly drawing out and confronting the flawed thinking head on (Donovan & Bransford, 2005).

Worked Examples With Self-Explanation

Worked examples—mathematics problems with worked-out solutions provided—offer an opportunity to call students' attention to common misconceptions. Some textbooks provide one, or a small number, of worked examples at the beginning of a problem set; students are simply asked to study a problem solution rather than solve a problem themselves. Assignments with interleaved worked examples, in contrast, embed the worked solutions throughout a problem set, alternating between examples and problems for students to solve. Positive effects of interleaving worked examples have been reported in a variety of courses (Clark & Mayer, 2003), including algebra (Sweller & Cooper, 1985). Replacing many of the problems in a practice session with examples of how to solve a problem leads to the same amount of procedural learning in less time (Clark & Mayer, 2003; Zhu & Simon, 1987), or increased learning and transfer of knowledge in the same amount of time (Paas, 1992). Worked examples of problem solutions are thought to be more efficient for learning new tasks because they reduce the load in working memory (compared with completing long strings of practice problems), thereby allowing students to learn the steps in problem solving (Sweller, 1999).

The value of worked examples can be enhanced with self-explanation prompts. Numerous empirical results have supported the notion that self-explanation is beneficial for learning (see Chi, 2000, for a review). When individuals self-explain, they integrate various pieces of knowledge (either from the instructional material, their own prior knowledge, or a combination of the two), generate inferences to fill gaps in their own knowledge, and make explicit the new knowledge and the connections that they've generated (Chi, 2000; Roy & Chi, 2005). High-ability students tend to spontaneously self-explain more often than low-ability students (Chi, Bassock, Lewis, Reimann, & Glaser, 1989), but students of any level who are prompted to self-explain learn more than those who do not self-explain (Chi, de Leeuw, Chiu, & Lavancher, 1994).

Empirical laboratory studies have shown that asking students to explain the errors in incorrect solutions, as well as the successful strategy in correct solutions, leads to greater learning than explaining correct solutions only (Durkin & Rittle-Johnson, 2009; Grosse & Renkl, 2004, 2007; Rittle-Johnson, 2006; Siegler, 2002; Siegler & Chen, 2008). These studies suggest that two of the main mechanisms by which incorrect examples improve learning are providing negative feedback and causing cognitive conflict. Negative feedback reduces the relative strength of incorrect strategies, which, when coupled with correct examples or instruction on correct procedures, is even more likely to cause procedural improvement. Cognitive conflict forces students to see the differences between the presented problem and others where a procedure does work, which in turn strengthens the likelihood of causing conceptual improvement. Recent work suggests that incorrect examples may be even more important than correct examples for promoting conceptual understanding (J. L. Booth, Lange, Koedinger, & Newton, 2013).

Incorrect examples can be used to target common misconceptions that make solving a particular type of problem difficult. For example, students may commonly use problem-solving strategies that produce right answers in some situations (e.g., combine two terms by adding the numbers involved; $4x + 3x$ is $7x$), yet harbor misconceptions (about the nature of variable vs. constant terms) that become apparent when they attempt to generalize this strategy to problems where it is not appropriate (e.g., $4x + 3$ is not $7x$). When students study and explain incorrect examples, they directly confront these faulty concepts. They are less likely to acquire or maintain misconceptions because they have identified what is wrong, and explained why it is wrong (Ohlsson, 1996; Siegler, 2002).

Many studies have established the benefits for procedural knowledge of worked examples (e.g., Sweller & Cooper, 1985; Zhu & Simon, 1987), and the benefits for conceptual understanding of self-explanation (e.g., Chi, 2000). However, few studies were conducted in classroom settings with real teachers teaching their own students. The studies in most publications consisted of single classroom lessons. Further, although this research-backed approach had been recommended for instructional use by the U.S. Department of Education (Pashler et al., 2007), worked examples are not now a common feature of mathematics textbooks or other classroom resources.

MAKING WORKED EXAMPLES WORK IN THE CLASSROOM

In the SERP–MSAN partnership, as in all SERP partnerships, once the problem identification and framing phase has produced a direction for the work, a smaller research and development team is formed with the appropriate expertise to execute the plan. With the identification of worked examples as a viable approach to improving Algebra 1 outcomes, Ken Koedinger and Julie Booth became lead researchers. A working group was formed that included a mathematics coordinator and two Algebra 1 teachers from each of the participating school districts. In early meetings, the group explored findings from the research literature, and compared Algebra 1 assignments from 11 MSAN classrooms. The typical assignment contained 10 to 12 problems to solve. Of the 128 items across all of the assignments, only six requested an explanation and only one provided a worked example of a problem solution. Similarities and differences in algebra textbooks, in mathematics instruction, and in culture across the districts, and opportunities and constraints relevant to conducting a research project were also explored. Math coordinators and participating teachers had come to feel that this research process was truly collaborative and that teachers were being listened to in the planning meetings. As a result, they succeeded in persuading teachers in all of the districts to engage in a random assignment study.

Developing the Assignments

The initial assignments with interleaved worked examples were drafted by the research team to target concepts and misconceptions that were identified as problematic for students based on the research literature, the prior findings of the research team, and the experience of the practitioner codevelopers. The practitioner codevelopers then reviewed, revised, and, in some cases, reshaped the assignments so that the language and level of challenge was appropriate to the student population. An initial bank of 24 Algebra I assignments was produced and tested for usability and feasibility with a subset of teachers across five MSAN districts during the 2008–2009 school year. Results of a pilot study with three classrooms of students ($n = 51$) using four assignments were extremely promising. Half of the randomly chosen students ($n = 26$) in each participating class completed four AlgebraByExample assignments over the course of their chapter on equation solving; the other half ($n = 25$) completed control assignments. Students also completed a pretest and posttest of conceptual and procedural knowledge. Results revealed that students who received AlgebraByExample assignments improved more than students who received control assignments. The effect was even more pronounced for minority students and for low-achieving students. Though an appreciable achievement gap was found at pretest (63% vs. 72% correct), no difference was found between groups at posttest (both Caucasian and minority students answered 72% of problems correctly; J. L. Booth et al., in revision).

Teachers who used the AlgebraByExample assignments indicated that the assignments prompted interesting classroom discussions about students' misconceptions; teachers also felt their students learned more from these assignments than from typically designed control assignments. Collaborating teachers who asked individual students to think aloud while working with an assignment also noted that the examples both caused students to confront their misconceptions and helped them figure out how to solve other problems in the assignment.

Focus-group teachers provided suggestions on how the approach could be more easily implemented and made more useful for improving student learning. First, the teachers indicated that the scope of individual assignments should be tightened so each covered content that would typically be taught within a single lesson. Teachers were also concerned that individual misconceptions or common errors were typically only covered once in the whole set of assignments. They suggested that it would be better to have a higher dosage of assignments on key topics, so that misconceptions and common errors could be targeted. Based on the student learning results and teachers' embrace of the assignments approach as a whole, the SERP team concluded that the approach was promising, but that further development of the materials was necessary before an efficacy study was warranted.

Refining the Assignments

In summer 2010, SERP was awarded a 3-year development grant from the Institute of Education Sciences, U.S. Department of Education, to further develop and test the assignments, and to explore the role of motivation as a mediator of impact. In year 1, the number of assignments increased from 24 to 42, and the number of items on the assignments shortened from 12 items to 6–8 items. The content of each assignment was narrowed to better fit within a single lesson plan. Pilot studies were conducted with three or four assignments done in class to determine feasibility, and to gather data from small samples of students. Following the single unit studies, nine major algebra topics were agreed upon by the researchers and practitioners, and the content within each unit was adjusted to fit a wider range of curricula. Specific items were also refined to better address critical misconceptions.

In 2011–2012, a series of double-unit studies containing six to 14 assignments were completed across six districts. Results indicated that with more assignments used, AlgebraByExample led to improved conceptual scores and equivalent procedural scores. In addition, non-Asian minority students benefited more from AlgebraByExample even when prior ability was controlled for, as measured by the pretests. In preparation for the year-long study, items were further refined and assignments were reformatted and compiled into a spiral-bound workbook to ensure ease of use throughout the school year; this also reduced the danger of data loss because teachers distributed and collected workbooks each time they were used.

Testing the Assignments

In the 2012–2013 school year, AlgebraByExample was tested in 28 classrooms taught by 12 teachers of nonhonors Algebra I from five MSAN school districts across five states. Individual classrooms were randomly assigned to ensure that each participating teacher had at least one treatment and one control classroom to control for teacher variability. One teacher (3 classrooms) did not complete the study; thus the data for those students were excluded from all analyses. The final sample consisted of 380 students (189 experimental, 191 control; 47% boys; 50% low socioeconomic status) in 25 classrooms (13 treatment classrooms; 12 control classrooms). The ethnicity breakdown was: 30% White, 39% Black, 18% Hispanic, 7% Asian, and 6% biracial. Students were classified as *underrepresented minority* (URM; Black, Hispanic, biracial) or non-URM students (White, Asian); 63% of the students were classified as URM. The entire study was conducted in a typical course setting, with testing done as part of normal classroom activities and assignments administered by teachers as routine class work.

MEASURES

Prior Knowledge

Students' prior knowledge was assessed at pretest using a paper-and-pencil test consisting of 81 items. The measure was designed based on the views of collaborating teachers about the prealgebraic knowledge they would ideally want their Algebra I students to have at the start of the year. The percentage of prior knowledge items answered correctly was computed for each student. Sample items can be found in Table 1.

Motivation

Student motivation was assessed at pretest with brief, age-adapted versions of two well-established measures in the achievement motivation literature: Interest (Elliot & Harackiewicz, 1996) and competence expectancy (Elliot & Church, 1997). Students provided responses on Likert-type scales from 1 (no, not at all) to 7 (yes, definitely), and scores for each measure were calculated according to previously established guidelines (Elliot & Church, 1997; Elliot & Harackiewicz, 1996).

Conceptual and Procedural Knowledge

We operationally define *conceptual knowledge* as an understanding of the core features in problems for a given topic, and *procedural knowledge* as the ability to carry out procedures to solve problems in that topic (e.g., J. L. Booth, 2011). Conceptual and procedural knowledge were assessed using a single paper-and-pencil test consisting of 66 items (41 conceptual, 25 procedural). The percentage of conceptual items answered correctly and the percentage of procedural items answered correctly were computed for each student. Sample items can be found in Table 1.

Standardized Test Items

Students were administered the paper-and-pencil test utilized by J. L. Booth, Barbieri, Eyer, and Paré-Blagoev (2014). The test consisted of 10 algebra-related released items taken from the five standardized tests used by the participating districts: *Ohio Achievement Test—Grade 8* (3 items; Ohio Department of Education, 2006); *Standards of Learning Test—Grade 8 Mathematics* (1

TABLE 1
Sample Test Items

Assessment	Sample Items	
Prior knowledge	State whether each of the following is equivalent to x + 4 – 2 + x: a. (x + 4) – (2 + x) Yes No b. 4 + x – 2 + x Yes No c. x + (4 – 2) + x Yes No d. x + 4 – x + 2 Yes No e. (x + 4) + (–2 + x) Yes No f. x + 4 + x – 2 Yes No g. x + 2(2 – 1) + x Yes No	Evaluate each expression for the values $x = 2$, $y = -3$, and $z = 4$. a) $5x + y \cdot z$ b) $x - 3(y + z)$ c) $\frac{x}{z} + 2y$
Conceptual posttest	State whether each of the following is true for the quadratic function $y = -x^2 - 2x + 3$ a. The axis of symmetry is $x = 1$ Yes No b. The vertex is a minimum Yes No c. The vertex is $(-1,4)$ Yes No d. The vertex is $(-1,0)$ Yes No e. The vertex is $(4,-1)$ Yes No	State how many terms the resulting expression will have: a. $2x(3 + 4x)$ 2 3 4 b. $(2x + 3)(4x - 1)$ 2 3 4 c. $(x + 2)^2$ 2 3 4 d. $(x + 4)(x - 4)$ 2 3 4 e. $2(3x^2 - 4x + 1)$ 2 3 4 f. $(2x^2 + 3)(4x - 1)$ 2 3 4
Procedural posttest	**Simplify each expression using only positive exponents.** a. $a^{-2}b^2c^0$ b. $(m^4k^2)^3$	**Factor each expression completely.** a. $3x^2 - 18x$ b. $-6x^4 - 3x^3 + 9x^2$
Standardized items	What is the slope of the line containing the points (–2, 5) and (1, –7)? a. –4 b. –2 c. 2 d. 4	Which is one value of the set of x that makes the following true? $7x + 3 > 17$ a. 0 b. 1 c. 2 d. 3

item; Virginia Department of Education, 2008), *Illinois Standards Achievement Test—Grade 8 Math* (3 items; Illinois State Board of Education, 2009), *Wisconsin Knowledge and Concepts Examination—Grade 8 Mathematics* (2 items; Wisconsin Department of Public Instruction, 2006), and the EXPLORE test (1 item; American College Test, 2014). For each student, the percentage of problems answered correctly was computed. Sample items can be found in Table 1.

Teacher Reports

At the end of the year, teachers were administered a survey about their experience in the AlgebraByExample study. In one item, they were asked about the frequency with which they reviewed study assignments in class. Teachers responded by selecting one of the following options: *0–20% of the time*; *20–40% of the time*; *40–60% of the time*; *60–80% of the time*; or *80–100% of the time*. Teacher responses were recoded into a 1 (*0-20%*) to 5 (*80–100%*) scale. In addition, teacher checklists were used to determine how many assignments were used in each of their classes.

Instructional Manipulation

Workbooks containing 42 assignments were provided to students in the beginning of the school year (see Table 2 for a list of assignment topics). Two versions of the workbook were created. One contained all of the treatment versions (worked example and self-explanation prompt) of the assignments, and the other contained all of the control versions of the assignments. Each control assignment contained 6–8 problems to solve that were isomorphic to those found in the relevant textbook chapters (the majority of assignments had eight items, however the quadratics assignments only had six, as each item typically takes longer to complete). Each treatment assignment contained the same 6–8 items, but worked examples were provided for the four left-hand items (in quadratics assignments, there were only three left-hand items). Most assignments contained two correct examples to explain and two incorrect examples to explain, and all assignments had at least one of each type of example. See Figures 1 and 2 for excerpts of a treatment and a corresponding control assignment. Assignment workbooks were collected at the end of the school year, and student work was coded to determine how many of the left-hand items and right-hand items on each assignment were sufficiently attempted by the student. In the AlgebraByExample assignments, left-hand items are worked examples and right-hand are traditional problems for students to solve on their own. Control assignments had only traditional problems. The number of assignments for which at least 75% of the left-hand items were

TABLE 2
Comprehensive List of AlgebraByExample Assignment Topics

Absolute value	Writing proportions	Multiplication and division
Combining like terms	Solving proportions	properties of exponents
Decimals	Writing expressions and equations	Power to power properties of exponents
Distributive property	from words	Zero and negative properties of exponents
Fractions	Solving systems of equations	Multiple properties of exponents
Order of operations	by graphing	Graphing tables of exponents
Analyzing answers to	Solving systems of equations by	Growth and growth rate
word problems	elimination with multiplication	Decay and decay rate
Graphing linear equations	Solving systems of equations	Decay and growth rate
Slope	by elimination with adding	Adding and subtracting polynomials
Slope-intercept form	and subtracting	Multiplying monomials
Writing equations in slope-	Solving systems of equations	by polynomials
intercept form	by substitution	Factoring binomials
Solving 1- and 2-step equations	Adding and subtracting inequalities	Multiplying binomials
Solving multi-step equations	Multiplying and dividing inequalities	Quadratic formula
Solving multi-step equations	Graphing inequalities	Solving quadratics by factoring
with fractions	Compound inequalities	Solving quadratics using square root
	Writing inequalities from words	Graphing quadratic functions

attempted and the number of assignments for which at least 75% of the right-hand items were attempted was computed for each student.

PROCEDURE

Soon after the school year started and before any assignments were administered, all classes were administered the prior-knowledge test and the motivation survey. During the school year, teachers administered study assignments when they were relevant to the content being covered in their classes; teachers were asked to make sure they used the same assignment topics in both treatment and control classrooms. Student assignment workbooks were collected at the end of the school year. At the end of the year, all classes were administered the conceptual posttest, procedural posttest, and the standardized-test item measure during a single class period. Deidentified demographic data were provided by the school districts.

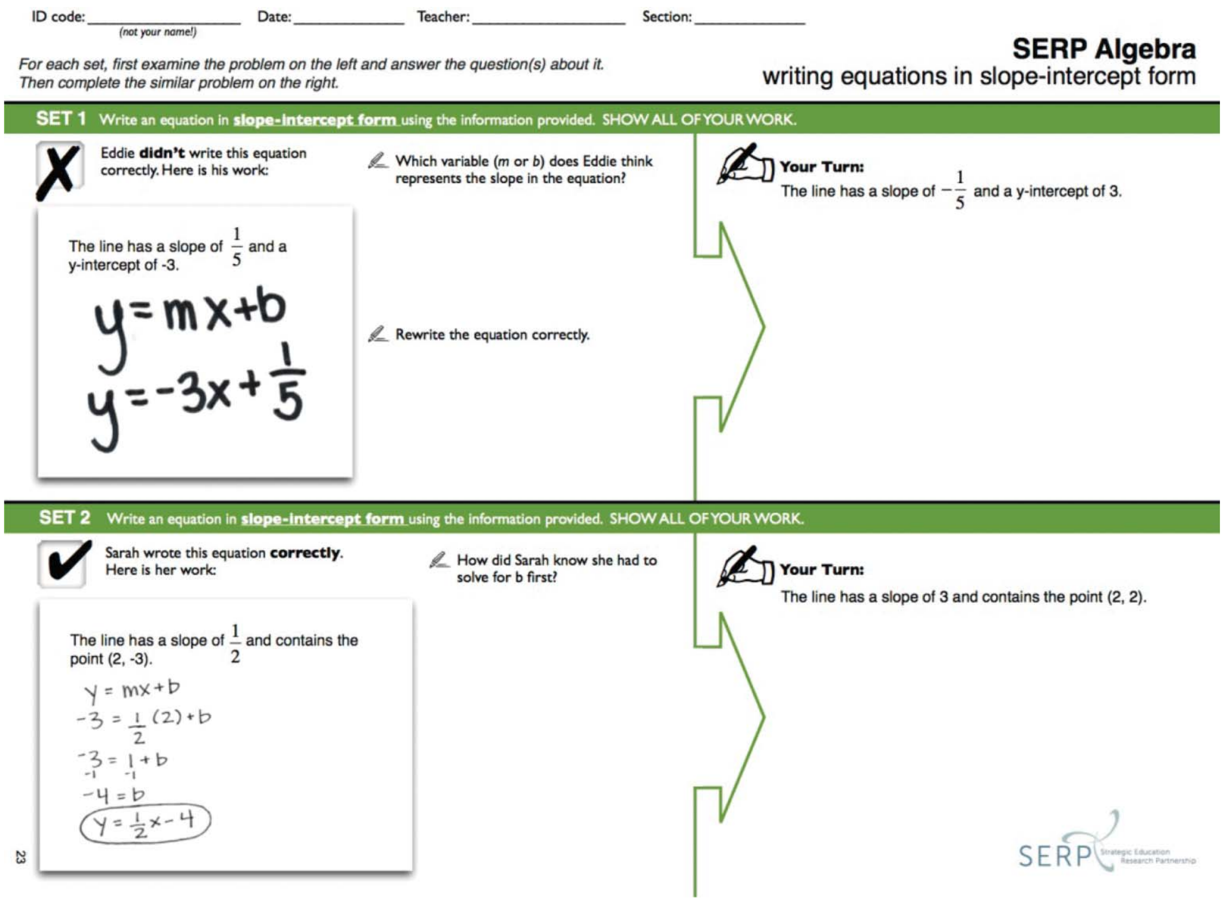
RESULTS

The research team conducted analyses in which students were nested in classrooms, with a sufficient number of clusters (25) to warrant the use of multilevel modeling (Maas & Hox, 2005). Thus, all analyses were conducted with hierarchical linear modeling (Raudenbush, Bryk, & Congdon, 2004).

All analyses used student data at Level 1 and classroom data at Level 2. Level 1 included the student's prior-knowledge scores and posttest conceptual, procedural, and standardized test item scores. Level 1 also included pretest interest and competence expectancy scores, the student's URM status (URM vs. non-URM), and the number of assignments on which students demonstrated sufficient effort on the left-hand items and the number on which they demonstrated sufficient effort on the right-hand items. Level 2 included whether or not the classroom was randomly assigned to the treatment condition, the proportion of the classroom that was comprised of URM students, the number of assignments the teacher used in the class, and the frequency with which the teacher reported reviewing the study assignments in class. Two students completed the posttest conceptual and procedural test, but not the posttest standardized items test. To include the maximum number of students in each analysis, students with missing data were excluded when running individual analyses. Descriptive statistics may be found in Table 3.

Effectiveness of the AlgebraByExample Intervention

To determine whether worked example assignments improved learning for algebra students, and whether there are differences in benefit based on student-level individual differences, we conducted a series of two-level hierarchical linear models with individual students nested in classrooms. For conceptual, procedural, and standardized item posttest scores, we first tested an empty model and determined that 55% of the variance in posttest conceptual scores, 41% of the variance in posttest procedural scores, and 44% of the variance in posttest standardized item scores was between classrooms, supporting the need for multilevel modeling. Because all



ID code: ________________ Date: ____________ Teacher: ________________ Section: ____________
(not your name!)

For each set, first examine the problem on the left and answer the question(s) about it. Then complete the similar problem on the right.

SERP Algebra
writing equations in slope-intercept form

SET 1 Write an equation in **slope-intercept form** using the information provided. SHOW ALL OF YOUR WORK.

Eddie **didn't** write this equation correctly. Here is his work:

The line has a slope of $\frac{1}{5}$ and a y-intercept of -3.

$y = mx + b$

$y = -3x + \frac{1}{5}$

Which variable (*m* or *b*) does Eddie think represents the slope in the equation?

Rewrite the equation correctly.

Your Turn:
The line has a slope of $-\frac{1}{5}$ and a y-intercept of 3.

SET 2 Write an equation in **slope-intercept form** using the information provided. SHOW ALL OF YOUR WORK.

Sarah wrote this equation **correctly**. Here is her work:

The line has a slope of $\frac{1}{2}$ and contains the point (2, -3).

$y = mx + b$

$-3 = \frac{1}{2}(2) + b$

$-3 = 1 + b$
$-1 \quad -1$

$-4 = b$

$y = \frac{1}{2}x - 4$

How did Sarah know she had to solve for b first?

Your Turn:
The line has a slope of 3 and contains the point (2, 2).

23

SERP Strategic Education Research Partnership

FIGURE 1 Excerpt from the treatment version of the "Writing Equations in Slope-Intercept Form" assignment.

ID code: ____________ Date: __________ Teacher: ______________ Section: __________
(not your name!)

SERP Algebra
writing equations in slope-intercept form

SET 1 Write an equation in **slope-intercept form** using the information provided. SHOW ALL OF YOUR WORK.

1a. The line has a slope of $\frac{1}{5}$ and a y-intercept of -3.

1b. The line has a slope of $-\frac{1}{5}$ and a y-intercept of 3.

SET 2 Write an equation in **slope-intercept form** using the information provided. SHOW ALL OF YOUR WORK.

2a. The line has a slope of $\frac{1}{2}$ and contains the point (2, -3).

2b. The line has a slope of 3 and contains the point (2, 2).

23

SERP

FIGURE 2 Excerpt from the control version of the "Writing Equations in Slope-Intercept Form" assignment.

TABLE 3
Descriptive Statistics for Level 1 and Level 2 Variables

	N	*Mean*	*Standard Deviation*	*Minimum*	*Maximum*
Student level variables: Level 1					
URM	380	0.63	0.48	0.00	1.00
Prior knowledge score	380	0.53	0.19	0.01	0.99
Pretest competence expectancy score	380	5.86	1.19	1.00	7.00
Pretest interest score	380	4.40	1.65	1.00	7.00
Assignment effort—Right-hand	380	18.80	11.09	0.00	40.00
Assignment effort—Left-hand	380	18.30	11.14	0.00	40.00
Posttest procedural score	380	0.25	0.24	0.00	1.00
Posttest conceptual score	380	0.50	0.22	0.00	0.98
Posttest standardized item score	378	0.49	0.34	0.00	1.00
Classroom level variables: Level 2					
Treatment	25	0.52	0.51	0.00	1.00
Proportion URM (class composition)	25	0.50	0.33	0.07	1.00
Assignments used	25	26.32	8.47	15.00	40.00
Reviewed	25	3.00	1.04	1.00	5.00

Note. URM = Underrepresented minority.

intraclass correlations were significant, we report the estimates for robust standard errors (Raudenbush & Bryk, 2002). The focal variables in the subsequent models were treatment (whether the students received the worked example assignments or not) and the interactions between treatment and prior knowledge and treatment and URM. Pretest scores on the interest and competence expectancy scales, effort on left-hand and right-hand items, as well as the percentage of URM per classroom, the number of assignments used in each classroom, and the frequency with which assignments were reviewed in the classroom are included in each model as control variables.

Conceptual Knowledge

As shown in Table 4, after controlling for interest and competence expectancy at pretest, assignment use, assignment effort, and URM status, model 1 indicated that students with higher prior knowledge scored higher on the posttest conceptual test; students in classrooms where the teacher reported reviewing the study assignments frequently also had higher posttest conceptual scores. Model 1 also yielded a trend toward a main effect of treatment, with students in the treatment group outperforming those in the control group. When the interactions between treatment and both prior knowledge and URM were included, model 2 yielded a significant main effect of treatment (see Figure 3) and a significant interaction between treatment and prior knowledge, indicating that the influence of condition on conceptual posttest scores varied by students' prior knowledge levels. As shown in Figure 4, the benefit of being in the treatment group was especially strong for students with lower prior knowledge. The Bayesian Information Criterion (BIC) decreased slightly from model 1 (−376) to model 2 (−377) indicating a slightly better fit of model 2 to the data.

TABLE 4
Predictors of Posttest Conceptual Scores

	Model 1 Main Effects Model	*Model 2 Interaction Model*
Fixed effects		
Student-level		
URM	−.01 (.02)	.01 (.02)
Classroom-level		
Intercept	.34** (.05)	.33** (.05)
Treatment	.06[†](.03)	.06* (.03)
Proportion URM (class composition)	−.12** (.04)	−.11** (.04)
Assignments used	.00 (.00)	.00 (.00)
Frequency of review	.04* (.02)	.04* (.02)
Treatment × Prior knowledge interaction	—	−.34** (.12)
Treatment × URM interaction	—	−.04 (.03)
Random effects		
Student-level		
Prior knowledge score[a]	.21* (.08)	.38** (.08)
Pretest interest score[a]	.00 (.01)	.00 (.01)
Pretest competence expectancy score[a]	−.01 (.01)	−.01 (.01)
Assignment effort—Right-hand[a]	.01 (.00)	.01 (.00)
Assignment effort—Left-hand[a]	.00 (.00)	.00 (.00)
Variance components		
Classroom-level	.0048	.0052
Student-level	.0131	.0128
Proportion reduction in variance (from Model 1)	—	.0191

Note. URM = Underrepresented minority. β (SE). — = not included in model.
[a]Predictor is grand-mean centered.
[†]$p < .10$; $*p < .05$; $**p < .01$.

Procedural Knowledge

Parallel analyses were conducted to understand predictors of posttest procedural scores. As shown in Table 5, after controlling for interest and competence expectancy at pretest, assignment use, assignment effort, and URM status, students with higher prior knowledge scored higher on the posttest procedural test; students in classrooms where the teacher reported reviewing the study assignments frequently also had higher posttest procedural scores. Model 1 also yielded a trend toward a main effect of treatment, with students in the treatment group outperforming those in the control group. Model 2 yielded a significant main effect of treatment (see Figure 5). The influence of treatment on procedural scores did not vary by students' prior knowledge or URM status. The BIC did not decrease from model 1 (−466) to model 2 (−461).

Standardized Test Items

Parallel analyses were conducted to understand predictors of performance on standardized test items. As shown in Table 6, after controlling for interest and competence expectancy at pretest,

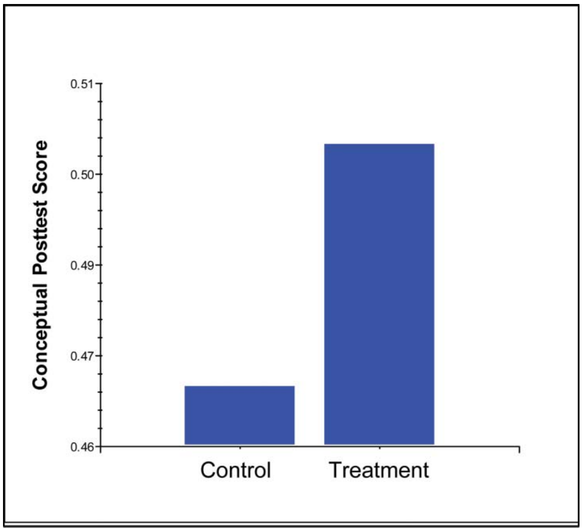


FIGURE 3 Effect of treatment on posttest conceptual scores.

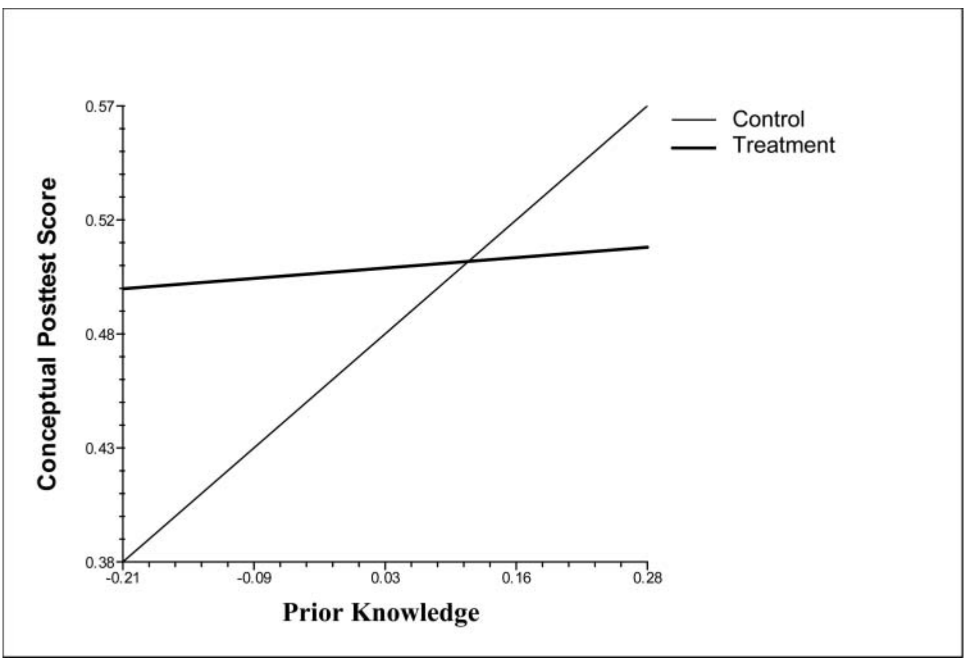


FIGURE 4 Interaction between prior knowledge and treatment on posttest conceptual scores.

TABLE 5
Predictors of Posttest Procedural Scores

	Model 1 Main Effects Model	*Model 2 Interaction Model*
Fixed effects		
Student-level		
URM	−.01 (.02)	.01 (.03)
Classroom-level		
Intercept	.11* (.05)	.10† (.05)
Treatment	.03† (.01)	.06* (.03)
Proportion URM (class composition)	−.07** (.02)	−.07** (.02)
Assignments used	.00 (.00)	.00 (.00)
Frequency of review	.03** (.01)	.03** (.01)
Treatment × Prior knowledge interaction	—	.06 (.09)
Treatment × URM interaction	—	−.03 (.04)
Random effects		
Student-level		
Prior knowledge score[a]	.40** (.05)	.37** (.06)
Pretest interest score[a]	.00 (.00)	.00 (.00)
Pretest competence expectancy score[a]	.00 (.00)	.00 (.00)
Assignment effort—Right-hand[a]	.00 (.00)	.00 (.00)
Assignment effort—Left-hand[a]	.00 (.00)	.00 (.00)
Variance components		
Classroom-level	.0028	.0029
Student-level	.0110	.0109
Proportion reduction in variance (from Model 1)	—	.0046

Note. URM = Underrepresented minority. β (SE). — = not included in model.
[a]Predictor is grand-mean centered.
†$p < .10$; *$p < .05$; **$p < .01$.

assignment use, assignment effort, and URM status, both model 1 and model 2 yielded a main effect of treatment on standardized test item scores, such that students in the treatment group tended to have higher standardized test item scores at posttest (see Figure 6). Both models also indicated that students with higher prior knowledge scored higher on the standardized test items, and students in classrooms where the teacher reported reviewing the study assignments frequently also had higher standardized test item scores. The influence of treatment on standardized test item scores did not vary by students' prior knowledge level or URM status. The BIC did not decrease from model 1 (24) to model 2 (28).

DISCUSSION

AlgebraByExample provides a boost in performance, with the greatest impact on students at the lower end of the performance distribution. On the researcher-designed assessment of conceptual understanding, treatment students in the lower half of the performance distribution outscored comparable control students by approximately 10 percentage points. Treatment students overall scored 7 percentage points higher on a test composed entirely of released items from the state standardized tests, and 5 percentage points higher on the conceptual posttest. Procedural posttest scores were also 4 percentage points higher in the

TABLE 6
Predictors of Posttest Standardized Item Scores

	Model 1 Main Effects Model	*Model 2 Interaction Model*
Fixed effects		
Student-level		
URM	−.05† (.03)	−.01 (.05)
Classroom-level		
Intercept	.45** (.07)	.43** (.08)
Treatment	.06* (.03)	.13* (.06)
Proportion URM (class composition)	−.16** (.03)	−.15** (.03)
Assignments used	−.01* (.00)	−.01* (.00)
Frequency of review	.06** (.02)	.06** (.02)
Treatment × Prior knowledge interaction	—	−.09 (.18)
Treatment × URM interaction	—	−.09 (.06)
Random effects		
Student-level		
Prior knowledge score[a]	.53** (.09)	.59** (.16)
Pretest interest score[a]	.00 (.01)	.00 (.01)
Pretest competence expectancy score[a]	−.02† (.01)	−.02† (.01)
Assignment effort—Right-hand[a]	.00 (.01)	.00 (.01)
Assignment effort—Left-hand[a]	.01 (.01)	.01 (.01)
Variance components		
Classroom-level	.0032	.0033
Student-level	.0406	.0400
Proportion reduction in variance (from Model 1)	–	.0145

Note. URM = Underrepresented minority. β (SE). — = not included in model.
[a]Predictor is grand-mean centered.
†$p < .10$; *$p < .05$; **$p < .01$.

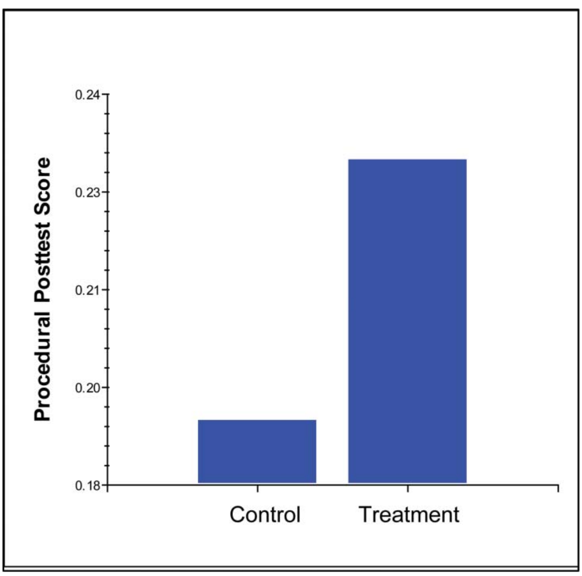


FIGURE 5 Effect of treatment on posttest procedural scores.

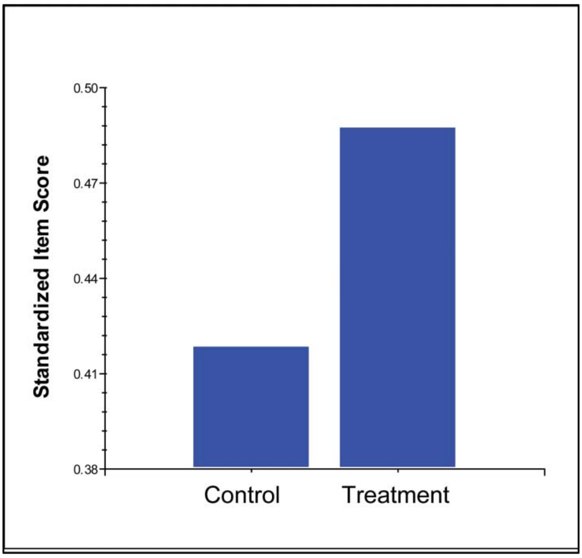


FIGURE 6 Effect of treatment on standardized item scores.

treatment group, even though control students had twice the practice solving problems on the assignments. Of equal significance, AlgebraByExample achieved these gains with an intervention that meets all the constraints imposed by the districts: it targets all students, it can be used with any existing Algebra 1 curriculum, and it is an asset that teachers can use with minimum disruption to their practice. Moreover, study teachers reported that students using AlgebraByExample required less teacher support to complete assignments than those using control assignments, and reported rethinking their own practice in response to students' positive experiences with worked examples.

The SERP–MSAN partnership departed from traditional education research studies in the extent to which the intervention was both shaped by the needs and the input of practitioners and tested with rigorous research methodologies, minimizing limitations from both practitioner and research perspectives. For teachers, limitations included constraints of the research study, such as not being able to use the assignments as homework (to minimize data loss), grade the assignments (except for completion), or use the AlgebraByExample assignments in all of their classrooms.

From a research perspective, the limitations of the study included a sample restricted to suburban or small urban areas and regular, nonhonors classrooms. The impact is not known for students in large urban districts, or in honors or remedial classrooms. In addition, although a within-teacher design controlled for teacher variables, it is possible that there was some contamination across classrooms.

The limitations of the study, however, are dwarfed by its benefits. The MSAN practitioners accomplished their goal of finding a practical intervention that improved African American and Latino students' understanding of algebra. Researchers also accomplished university goals, including tenure-related research and graduate student dissertations. And the SERP organization accomplished its goals of generating knowledge and products that are of value to the field of education more broadly. The assignments are not only accessible freely (math.serpmedia.org/algebra_by_example/), but they can be used without resource-intensive professional development or costly changes in curriculum.

Although the SERP-MSAN partnership was able to successfully navigate around obstacles to partnership success, it is worth noting that current funding mechanisms are particularly challenging in the long delays imposed by application and review schedules. Because MSAN is a partnership of many districts, the changing circumstances that led to withdrawal from the collaboration by some districts was successfully resolved when other districts joined the study, an affordance of the MSAN alliance that is unique. Although opportunities for funding research–practice partnerships are expanding, timing is likely to remain problematic unless expedited review and opportunities to respond quickly with revisions are put in place for partnership work that is already in motion.

ACKNOWLEDGMENTS

We thank all the researchers who helped frame this work, the Minority Student Achievement Network and Executive Director Madeline Hafner, and district leaders, teachers, and students from Ann Arbor Public Schools, Arlington Public Schools, Chapel Hill-Carrboro City Schools, Green Bay Public Schools, Evanston Township High School District 202, Evanston/Skokie School District 65, Madison Metropolitan School District, and Shaker Heights Public Schools.

FUNDING

Funding for this research was provided by the Goldman Sachs Foundation, and by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A100150 to the Strategic Education Research Partnership. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

REFERENCES

American College Test. (2014). *EXPLORE*. Iowa City, IA: Author.

Baroody, A., & Ginsburg, H. (1983). The effects of instruction on children's understanding of the equals sign. *The Elementary School Journal*, *84*, 199–212.

Booth, J. L. (2011). Why can't students get the concept of math? *Perspectives on Language and Literacy, 37*, 31–35.

Booth, J. L., Barbieri, C., Eyer, F., & Paré-Blagoev, E. J. (2014). Persistent and pernicious errors in algebraic problem solving. *Journal of Problem Solving, 7*, 10–23.

Booth, J. L., & Koedinger, K. R. (2008). Key misconceptions in algebraic problem solving. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Cognitive Science Society* (pp. 571–576). Austin, TX: Cognitive Science Society.

Booth, J. L., Koedinger, K. R., & Siegler, R. S. (2007, August). *The effect of prior conceptual knowledge on procedural performance and learning in algebra*. Poster presented at the 29th annual meeting of the Cognitive Science Society, Nashville, TN.

Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction*, *25*, 24–34. doi: 10.1016/j.learninstruc.2012.11.002

Booth, J. L., Oyer, M. H., Paré-Blagoev, E. J., Elliot, A., Barbieri, C., Augustine, A. A., & Koedinger, K. R. (in revision). Learning algebra by example in real-world classrooms. *Journal of Research on Educational Effectiveness*.

Booth, L. R. (1984). *Algebra: Children's strategies and errors*. Windsor, UK: NFER-Nelson.

Brown, D. E. (1992). Using examples and analogies to remediate misconceptions in physics: Factors influencing conceptual change. *Journal of Research in Science Teaching*, *29*, 17–34. doi: 10.1002/tea.3660290104

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In Glaser, R. (Ed.), *Advances in instructional psychology* (pp. 161–238). Mahwah, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H., Bassock, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145–182. doi: 10.1207/s15516709cog1302_1

Chi, M. T. H., de Leeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439–477. doi: 10.1207/s15516709cog1803_3

Chiu, M. H., & Liu, J. W. (2004). Promoting fourth graders' conceptual change of their understanding of electric current via multiple analogies. *Journal of Research in Science Teaching*, *42*, 429–464. doi: 10.1002/tea.20062

Clark, R. C., & Mayer, R. E. (2003). *e-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. San Francisco, CA: Jossey-Bass. doi: 10.1002/9781118255971

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, *32* (6), 3–12.

Donovan, M. S., & Bransford, J. D. (Eds.). (2005). *How students learn: History, mathematics, and science in the classroom*. Washington, DC: National Academy Press.

Durkin, K. L. & Rittle-Johnson, B. (2009, April). *Comparison of correct and incorrect examples when learning decimal fractions*. Poster presented at the annual meeting of the Society for Research in Child Development, Denver, CO.

Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, *72* (1), 218–232. doi: 10.1037/0022-3514.72.1.218

Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology*, *70*, 968–980. doi: 10.1037/0022-3514.70.3.461

Grosse, C. S. & Renkl, A. (2004). Learning from worked examples: What happens if errors are included? In P. Gerjets, J. Elen, R. Joiner, & P. Kirschner (Eds.), *Instructional design for effective and enjoyable computer-supported learning* (pp. 356–364). Tübingen, Germany: Knowledge Media Research Center.

Grosse, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning & Instruction*, *17*, 617–634. doi: 10.1016/j.learninstruc.2007.09.008

Harter, S., & Connell, J. P. (1984). A model of the relationship among children's academic achievement and their self-perceptions of competence, control, and motivational orientation. In J. Nichols (Ed.), *The development of achievement motivation* (pp. 219–250). Greenwich, CT: JAI Press.

Hemphill, F. C., & Vanneman, A. (2011). *Achievement gaps: How Hispanic and White students in public schools perform in mathematics and reading on the National Assessment of Educational Progress (NCES 2011-459)*. Washington, DC: U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences. doi: 10.1037/e595272011-001

Illinois State Board of Education. (2009). *Illinois standards achievement test: Grade 8 math*. Springfield, IL: Author.

Kendeou, P., & van den Broek, P. (2005). The effects of readers' misconceptions on comprehension of scientific text. *Journal of Educational Psychology*, *97*, 235–245. doi: 10.1037/0022-0663.97.2.235

Kennedy, M. (2005). *Inside teaching: How classroom life undermines reform*. Cambridge, MA: Harvard University Press. doi: 10.5860/CHOICE.43-2333

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics*, *12*, 317–326. doi: 10.1007/BF00311062

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education*, *37*, 297–312.

Küchemann, D. (1978). Children's understanding of numerical variables. *Mathematics in School*, *7*, 23–26.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1* (3), 86–92. doi: 10.1027/1614-1881.1.3.86

Ohio Department of Education. (2006). *Ohio achievement tests (8) mathematics student test booklet*. Columbus, OH: Author.

Ohlsson, S. (1996). Learning from error and the design of task environments. *International Journal of Educational Research*, *25* (5), 419–448. doi: 10.1016/S0883-0355(97)81236-0

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*, 429–434. doi: 10.1037/0022-0663.84.4.429

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning (NCER 2007–2004)*. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. doi: 10.1037/e607972011-001

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2004). *HLM (Version 6)* [Computer software]. Lincolnwood, IL: Scientific Software International.

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, *77*, 1–29. doi: 10.1111/j.1467-8624.2006.00852.x

Roy, M., & Chi, M. T. H. (2005). Self-explanation in a multi-media context. In R. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 271–286). New York, NY: Cambridge Press. Retrieved from http://dx.doi.org/10.1017/CBO9780511816819.018

Siegler, R. S. (2002). Microgenetic studies of self-explanations. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 31–58). New York, NY: Cambridge University Press. Retrieved from http://dx.doi.org/10.1017/CBO9780511489709.002

Siegler, R. S., & Chen, Z. (2008). Differentiation and integration: Guiding principles for analyzing cognitive change. *Developmental Science*, *11*, 433–448. doi: 10.1111/j.1467-7687.2008.00689.x

Steele, C. M. & Aronson, J. (1995). Stereotype vulnerability and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*, 797–811. doi: 10.1037/0022-3514.69.5.797

Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Washington, DC: Brookings Institution. doi: 10.2307/40253438

Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Australia: ACER Press.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, *2*, 59–89. doi: 10.1207/s1532690xci0201_3

Vanneman, A., Hamilton, L., Anderson, J. B., & Rahman, T. (2009). *Achievement gaps: How Black and White students in public schools perform in mathematics and reading on the National Assessment of Educational Progress (NCES 2009-45)*. Washington, DC: U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences. doi:10.1037/e566412009-001

Virginia Department of Education. (2008). *Standards of learning test: Grade 8 mathematics*. Richmond, VA: Author.

Vlassis, J. (2004). Making sense of the minus sign or becoming flexible in 'negativity.' *Learning and Instruction*, *14*, 469–484. doi: 10.1016/j.learninstruc.2004.06.012

Warren, E. (2003). The role of arithmetic structure in the transition from arithmetic to algebra. *Mathematics Education Research Journal*, *15*, 122–137. doi: 10.1007/BF03217374

Wisconsin Department of Public Instruction. (2006). *Wisconsin knowledge and concepts examinations criterion-referenced test: Guide to grade 8 released item books in reading and mathematics*. Monterey, CA: Wisconsin Department of Public Instruction & CTB/McGraw-Hill.

Zhu, X., & Simon, H. A. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction*, *4*, 137–166. doi: 10.1207/s1532690xci0403_1