

Journal of Research on Educational Effectiveness

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/uree20>

Learning Algebra by Example in Real-World Classrooms

Julie Booth^a, Melissa Oyer^a, E. Juliana Paré-Blagoev^b, Andrew J. Elliot^c, Christina Barbieri^a, Adam Augustine^c & Kenneth R. Koedinger^d

^a Temple University, Philadelphia, Pennsylvania, USA

^b Strategic Education Research Partnership, Washington, DC, USA

^c University of Rochester, Rochester, New York, USA

^d Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Accepted author version posted online: 28 Aug 2015.



[Click for updates](#)

To cite this article: Julie Booth, Melissa Oyer, E. Juliana Paré-Blagoev, Andrew J. Elliot, Christina Barbieri, Adam Augustine & Kenneth R. Koedinger (2015): Learning Algebra by Example in Real-World Classrooms, Journal of Research on Educational Effectiveness, DOI: [10.1080/19345747.2015.1055636](https://doi.org/10.1080/19345747.2015.1055636)

To link to this article: <http://dx.doi.org/10.1080/19345747.2015.1055636>

Disclaimer: This is a version of an unedited manuscript that has been accepted for publication. As a service to authors and researchers we are providing this version of the accepted manuscript (AM). Copyediting, typesetting, and review of the resulting proof will be undertaken on this manuscript before final publication of the Version of Record (VoR). During production and pre-press, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal relate to this version also.

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any

Learning Algebra by Example in Real-World Classrooms

Julie Booth and Melissa Oyer

Temple University, Philadelphia, Pennsylvania, USA

E. Juliana Paré-Blagoev

Strategic Education Research Partnership, Washington, DC, USA

Andrew J. Elliot

University of Rochester, Rochester, New York, USA

Christina Barbieri

Temple University, Philadelphia, Pennsylvania, USA

Adam Augustine

University of Rochester, Rochester, New York, USA

Kenneth R. Koedinger

Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Abstract: Math and science textbook chapters invariably supply students with sets of problems to solve, but this widely used approach is not optimal for learning; instead, more effective learning

can be achieved when many problems to solve are replaced with correct and incorrect worked examples for students to study and explain. In the present study, the worked example approach is implemented and rigorously tested in the natural context of a functioning course. In Experiment 1, a randomized controlled study in ethnically diverse Algebra classrooms demonstrates that embedded worked examples can improve student achievement. In Experiment 2, a larger randomized controlled study demonstrated that improvement in posttest scores as a result of the assignments varies based on students' prior knowledge; students with low prior knowledge tend to improve more than higher knowledge peers.

Keywords: Worked examples, prior knowledge, algebra learning

Address correspondence to Julie Booth. E-mail: julie.booth@temple.edu

Open any math or science textbook and you'll see problem sets that demonstrate reliance on a "practice-makes-perfect" approach, wherein students are first taught content and then provided with problems they are required to solve. The IES Cognition Practice Guide (Pashler et al., 2007) recommended that teachers move away from this ubiquitous method and instead provide their students with paired sets that include a worked-out example of a problem solution followed by a matched problem to solve. This recommendation is grounded in 35+ years of mostly laboratory studies showing the learning advantages of such a "worked example" approach. That the recommendation suggested teachers spend the time to create their own examples reflects the reality that there is a "practice gap" -- there are few ready-made materials for teachers to use that instantiate the worked examples approach. Further, there is a "research gap" -- it is not known whether using worked examples in classroom settings over longer time periods have the same or different effects as those found in brief experiments conducted in laboratory settings which are less complex both socially and pedagogically. Here, we describe results from two studies exploring student outcomes and individual student differences when worked examples are instantiated into classroom-friendly assignments and used as a part of real teachers' regular instruction over several weeks.

The worked example effect: A brief review

In 1985, Sweller and Cooper reported the then surprising finding that instead of having a learner practice solving problems, they learn more or faster if they spend half their time studying examples of how to solve problems and the other half practicing on their own. Since these

seminal studies, further laboratory studies have been conducted establishing the potency of this worked example approach across multiple STEM domains, including algebra (Carroll, 1994; Cooper & Sweller, 1987), geometry (Paas & Van Merriënboer, 1994; Tarmizi & Sweller, 1988), physics (Hausmann & VanLehn, 2007; Ward & Sweller, 1990), probability (Catrambone, 1998; Große & Renkl, 2007; Renkl, 2002), and chemistry (McLaren, Lim, Gagnon, Yaron, & Koedinger, 2006). This approach has also been tested with participants from a variety of different grade levels: Elementary school (Rittle-Johnson, 2006; Schorr, Gerjets, Scheiter, & Laouris, 2002; Siegler & Chen, 2008); middle school (Zhu & Simon, 1987); high school (Atkinson et al., 2003; Cooper & Sweller, 1987), and adulthood (Catrambone & Yuasa, 2006; Moreno, Reisslein, & Ozogul, 2009). Collectively, these studies establish that the worked example approach can be effective in laboratory settings and provide strong, empirically guided recommendations for how to design worked example activities.

One view is that worked examples improve student learning by reducing procedural demands on working memory, thereby increasing the cognitive bandwidth for conceptual learning and reducing the chances that students will acquire incorrect or poorly generalizable procedural habits (Sweller, 1999; Zhu & Simon, 1987). Some scholars have combined the use of worked examples with other methods to further enhance learning. For instance, sometimes worked examples are combined with a requirement that students respond to targeted self-explanation prompts about specific features of the example (e.g., Aleven & Koedinger, 2007; Hilbert, Renkl, Kessler, & Reiss, 2008). This requirement may act to facilitate integration of new information with prior knowledge and force the learner to make their new knowledge explicit (Chi, 2000; Roy & Chi, 2005). Another variation to the worked example approach is to reduce

intra-example variability by holding surface features of the example and practice problem constant. This may refocus students' attention on deep structures of the problems (Ben-Zeev & Start, 2001; Chang, 2006; Zhu & Simon, 1987). Some studies have also explored how incorrect vs. correct examples (marked as such) may differently influence learning; results suggest that when incorrect examples are used, students are less likely to acquire or maintain the demonstrated incorrect strategies (Siegler, 2002; Ohlsson, 1996). Related work also suggests that incorrect examples in particular can deepen students conceptual understanding -- not just improve their capacities to select and apply correct problem solving approaches (Booth, Lange, Koedinger, & Newton, 2013).

Why test worked examples in real-world classrooms?

With all of the accumulated evidence documenting and supporting the use of worked examples, it would be tempting to assume that the work is done, and that it is clear that this practice would produce appreciable gains in real-world educational settings. However, there are two primary reasons this conclusion cannot be drawn.

First, despite the wealth of studies showing that worked examples are effective in laboratory settings, there is some counterevidence suggesting that it does not always work. For example, Renkl (1997) demonstrated that just providing students with worked examples is not sufficient to promote learning; students must actively and appropriately engage with the examples. When left to their own devices, learners may fail to notice the key information depicted in worked examples, or even focus on irrelevant information (Ross, 1989). There is also evidence that students attempt to translate the strategies superficially when solving other problems and may have difficulty in generalizing these strategies to problems that are not

isomorphic to the example (Catrambone & Holyoak, 1989). In these cases, worked examples would not be said to have improved student learning. Further, there are other studies that call into question whether worked examples are beneficial for everyone or just some learners. For example, the level of the learner's prior knowledge has been shown to affect the degree to which worked examples are useful; typically, students with low prior knowledge are thought to improve more from worked examples than those with high prior knowledge (Kalyuga, Chandler, & Sweller, 2001). It has also been suggested that the use of worked examples should be faded out over time for optimal learning, such that learners are provided with fewer examples and more problem solving practice as they gain experience (Renkl, Atkinson, Maier, & Staley, 2002).

Second, it is not necessarily the case that interventions established in laboratory settings automatically translate to gains in real-world educational settings. Real-world classrooms are inherently less controlled and more contextual than university laboratories, and many strong laboratory findings fail to translate to effects in real-world classrooms. For example, although an abundance of laboratory findings exist supporting the multimedia principle -- that adding a relevant picture or diagram to text is more beneficial than text alone (Mayer, 1989; 2005) -- a straightforward application of this principle in large, real-world Chemistry classrooms did not show stronger outcomes from diagrams when compared with text alone (Davenport, Klahr, & Koedinger, 2007). Similarly, although the use of educational technology is widely purported to increase student scores and learning gains in mathematics and reading (Kulik, 1994; Murphy, Penuel, Means, Korbak, & Whaley, 2001), a large-scale randomized-controlled trial testing well-respected software for first grade reading, fourth grade reading, sixth grade math, and algebra yielded null results (Dynarski et al., 2007).

Given the complexity of instructional design (Koedinger, Booth, & Klahr, 2013), perhaps it shouldn't be surprising that laboratory gains do not automatically translate to the classroom. In fact, some educational researchers are quite skeptical of the value of laboratory research for educational practice, because "learning, cognition, knowing, and context are irreducibly co-constituted and cannot be treated as isolated entities or processes" (Barab & Squire, 2004, p. 1). These authors elaborate, "if one believes that context matters in terms of learning and cognition, research paradigms that simply examine these processes as isolated variables within laboratory ... will necessarily lead to an incomplete understanding of their relevance (Brown, 1992)" (p. 1). Certainly, the variety of academic material to which the worked example principle has been applied enhances its external validity. However, if issues of classroom implementation and context have not been adequately explored, the worked example principle has *not* been sufficiently tested.

Several studies investigating the worked example effect have been conducted in school settings. However, much of this work has been implemented in computer-based environments that provide individualized, real-time feedback, which is not possible or practical in traditional classrooms. For instance, some studies implemented this type of feedback during students' problem solving practice (Kalyuga, Chandler, Tuovinen, & Sweller, 2001; Kim, Weitz, Heffernan, & Krach, 2009; Paas, 1992; Schwonke et al., 2009), and feedback is sometimes given on students' explanations of the examples themselves (Booth, Lange, Koedinger, & Newton, 2013). An additional classroom-based study implemented typical example-problem pairs, but did not use a control group (Zhu & Simon, 1987). The two experimental studies conducted in traditional classrooms (Carroll, 1994; Ward & Sweller, 1990) occurred over very short periods of

time -- pretest, example study, and posttest all took place within two days, a timeframe that is not common in typical classroom activities and ignores much of the complexity inherent in real-world classrooms. Thus, it remains unclear whether the performance improvements due to worked examples that have been identified in the laboratory emerge in uncontrolled, traditional classroom settings.

Using Algebra as an optimal test-bed for the worked example approach

In the present study, we test the effectiveness of incorporating correct and incorrect examples into classroom assignments to improve the learning of concepts and procedures in Algebra I (late middle/early high school Algebra). Algebra is an excellent test-bed for application of these instructional principles. The subject can be particularly challenging not only because it introduces more abstract representations and more complex relationships between quantities, but also because it can magnify the misconceptions that have their roots in earlier instruction. A variety of particularly problematic misconceptions typically plague beginning algebra students, including believing that the equals sign is an indicator of operations to be performed (Baroody & Ginsburg, 1983; Kieran, 1981), that negative signs represent only the subtraction operation and do not modify terms (Vlassis, 2004), and that variables cannot take on multiple values (Booth, 1984; Knuth, Stephens, McNeil, & Alibali, 2006; Küchemann, 1978). Not surprisingly, such misconceptions have been shown to affect students' success in problem solving and hinder their learning of new material (Booth & Koedinger, 2008); unfortunately, many of these misconceptions persist even after classroom instruction (Booth, Koedinger, & Siegler, 2007; Vlassis, 2004).

Algebra I is also an important target because if students are unsuccessful in this course, they will be excluded from higher-level mathematics and science courses (Department of Education, 1997). The current work was carried out in the context of an ongoing research and development collaboration with the Minority Student Achievement Network (MSAN), a set of inner-ring suburban districts that have witnessed widening achievement gaps as their schools have seen increases in the diversity of their schools. Though the MSAN districts are especially concerned about their extant achievement gaps in algebra between underrepresented minority and non-minority students, they recognized that explicit interventions targeting minority students can highlight the historic poor performance of the targeted groups, resulting in reduced performance through stereotype threat (Steele, 1992; Steele & Aronson, 1995). Based on such research and their practical experience, the MSAN districts were adamant that the worked examples intervention aim to improve learning for all students, not just that of minority students. To address these concerns, the design of our intervention assignments, collectively “AlgebraByExample”, includes two elements that aim to provide subtle supports for all students, but which may be especially beneficial for disadvantaged students. First, the inclusion of incorrect examples is intended to help students realize that questions, mistakes, and incorrect or incomplete thoughts are not to be suppressed, as they may have implicitly been taught to do in traditional classroom activities (Ladson-Billings, 1995). Instead, errors are to be seen as a valuable component of the learning process; one that any student can make use of. Second, the strategic inclusion of student names from diverse cultures along with an equal distribution of correct and incorrect answers being attributed to boys vs. girls and students of varying cultural backgrounds aims to send the message that anyone can succeed or fail at this task, and everyone

can learn from mistakes. Consistent with these efforts, in the present studies, we examine overall improvements as a result of the assignments, as well as differential improvements from working with these assignments based on two factors highly relevant to these districts with prevalent achievement gaps: students' prior knowledge (Experiments 1 and 2) and school-level socioeconomic status (Experiment 2).

The Present Studies

In a series of two experiments, we extend the previous body of work by implementing worked examples in traditional Algebra 1 classrooms over the course of an entire unit (approximately 3-4 weeks of instructional time). Experiment 1 has two main purposes. First, it aims to explore, with a small sample, whether example-based assignments are more effective than typical problem solving practice when implemented in a traditional classroom setting. Results from numerous laboratory studies support the hypothesis that they will be effective, however, laboratory results do not always translate to appreciable effects in the classroom (Davenport et al., 2007).

Second, consistent with MSAN interests, we aim to determine whether differences in the effectiveness of the example-based assignments emerge for students based on their prior knowledge of the content. On one hand, the work of Kalyuga and colleagues (2001) suggests that students with low amounts of prior knowledge may experience the greatest improvement from studying worked examples; however, the work of Große & Renkl (2007) suggests that lower achieving students may not improve as much with worked examples when incorrect examples are included, especially if students must locate the error. Here we explore whether prior

knowledge impacts improvement when both correct and incorrect examples are included, but both are clearly marked and the error is highlighted.

Experiment 1

Method

Participants. 56 high school students (27 M, 29 F; mean age = 14.3 years, SD = .99 years) in three Algebra I classrooms from two MSAN districts in the mid-Atlantic region participated. ; All classrooms used the traditional Algebra 1 curricula and supplemented it with the AlgebraByExample assignments designed by our partnership. Two teachers volunteered their classes to participate in this intensive, 3-4 week experiment during their unit on solving linear equations. Five students (4M, 1F; 4 Caucasian, 1 Asian¹) were excluded from analyses because they took the pretest but were not present on the day the posttest was administered. The ethnicity breakdown of final sample was as follows: 47% Caucasian, 33% Black, 6% Hispanic, 8% Asian, and 6% of mixed race. Using free or reduced lunch eligibility as a proxy, 24% of the participating students were categorized as economically disadvantaged.

Random assignment was conducted at the level of the individual student. Approximately half of the students in each participating class were randomly assigned to the treatment group, in which they received the worked examples assignments designed for the study. The other half was randomly assigned to the control group, in which they received an alternate version of the assignments that contained the same types of problems, but no examples or self-explanation

¹ Chi squared analyses revealed no significant differences in demographics between the students who did not take the posttest and the rest of the sample. However, there was a trend towards a difference between pretest scores ($t[54] = -1.72, p = .09$) with excluded students ($M = .7857$) demonstrating higher prior knowledge than the rest of the sample ($M = .6849$). Thus, the final sample may have lower prior knowledge than the average for the population.

prompts (see Figure 1). The final sample included 26 students in the treatment condition and 25 students in the control condition.

Instructional manipulation. Researchers, teachers, and math administrators collaboratively developed both the worked example type and corresponding control assignments as part of SERP-MSAN partnership activities; all decisions about creation and ordering of problems and/or examples for each assignment were vetted and improved by practitioners. Four topics were covered in this study: Simple equations, Multi-step equations, Writing expressions and equations from words, and Finding reasonable answers to word problems. Each control assignment contained 10-12 problems to solve that were isomorphic to those found in the relevant textbook chapters. Each worked example assignment also contained 10-12 items, with base problems that were isomorphic to those in the control assignments: 2-3 correct examples to explain, 2-3 incorrect examples to explain, and 5-6 problems to solve. See Figure 1 for excerpts of an example-based and corresponding control assignment.

Measures. We operationally define procedural knowledge as knowing how to carry out a task, and conceptual knowledge as understanding the meaning of the features in the task, as well as why those features make a given procedure appropriate (e.g., Booth, 2011). We assume that both conceptual and procedural knowledge are necessary to do well in Algebra; this assumption is in accord with those of the practitioners involved in the project, as well as recommendations of experts (National Mathematics Advisory Panel, 2008).

Students were administered a content assessment as a pretest and posttest. This test was comprised of four procedural knowledge items, which required students to carry out procedures to solve problems; all four problems were isomorphic to problems in the assignments, which

were representative of the types of problems found in Algebra I textbooks and taught in Algebra I courses. The content test also included 22 conceptual knowledge items, which measured understanding of crucial concepts from the four assignments (e.g., the meaning of the equals sign, the significance of negatives in terms, the identification of like terms, etc.; Booth & Koedinger, 2008; Kieran, 1981; Knuth et al., 2006; Vlassis, 2004). Sample content test items may be found in Figure 2. Cronbach's alpha indicated that the internal consistency of the measure was sufficient ($\alpha = .70$). The percentage of items answered correctly was computed for each student at pretest and at posttest. Pretest scores did not differ for students in each condition ($M_{\text{treatment}} = .68$, $M_{\text{control}} = .69$, $t(49) = .353$, $p = .73$). We also conducted a median split on pretest scores, such that students with pretest scores below the median were coded as having low prior knowledge ($N=26$) and those above the median were coded as having high prior knowledge ($N=25$); there was no difference in the distribution of low and high prior knowledge students in each condition ($\chi^2[N=51] = .17$, $p = .68$).

Procedure. Before beginning their chapter on linear equations, all students were given identical in-class, paper and pencil pretests of their content knowledge of solving equations and algebraic word problems; no time limit was given, but the test took approximately 25 minutes to complete. During their equation-solving chapter, teachers taught the material at their usual pace and with their typical instructional methods and assignments. However, when they reached a particular topic that was covered in one of the four study assignments, students completed the worksheets on their own, as an in-class activity. Teachers distributed the randomly assigned version of the assignment to each student in their class, and provided time for independent practice. Students in the worked examples group received the AlgebraByExample versions of all

four assignments over the course of the study; students in the control group received four control assignments. As this was a within-class design, all other class experiences, activities, and assignments did not differ for experimental and control students. In order to maintain ecological validity (which may be low in highly controlled laboratory studies), teachers were permitted to discuss problem-solving items on the worksheets with their students as they saw fit. However, in order to maintain fidelity of implementation, teachers were asked not to discuss the worked examples from the AlgebraByExample assignments in class, so that the control students would not be exposed to the critical feature of the intervention. After completing the linear equations chapter, students were given the same content assessment again as a posttest. All study activities took place during regular class time.

Results

To assess the effectiveness of these assignments for the sample as a whole, as well as examine individual differences in their effectiveness, we conducted a 2 (Condition: Treatment vs. Control) x 2 (Prior knowledge: High vs. Low) ANOVA on posttest scores. Means and SDs may be found in Table 1. The analysis yielded a significant main effect of condition ($F[1,47] = 5.34, p = .03, \eta_p^2 = .10$); students in the treatment group outscored students in the control group. There was also a significant main effect of prior knowledge ($F[1,47] = 7.60, p = .008, \eta_p^2 = .14$), with high prior knowledge students outscoring low prior knowledge students on the posttest. The interaction between treatment and prior knowledge was not significant ($F[1,47] = 0.03, p = .86, \eta_p^2 = .001$).

Experiment 2

The results from Experiment 1 indicate that assignments with correct and incorrect examples to explain and problems to solve can indeed be more beneficial than standard, problem-based assignments when used in real world classrooms. This provides ecologically valid support for the lab-proven learning gains seen with worked examples (e.g., Sweller, 1999) and self-explanation (Roy & Chi, 2005). We were surprised to find no individual differences in improvements from the worked example assignments based on students' prior knowledge: Both low- and high prior knowledge students performed better after using the example-based assignments. However, as the participating students were all in the Fall of their Algebra I class and were learning the material for the first time, perhaps the entire sample could be considered novice learners; prior research would suggest that such a population of students would be especially likely to improve after studying examples (Kalyuga et al., 2001a; 2001b).

One strange finding in Experiment 1 is discovered by examining the mean scores at pretest and posttest for the individual groups in Table 2. Note specifically that low-achieving students' scores improved greatly in the treatment condition, but high-achieving students' scores actually decreased in the control condition. The result of this is that it appears all students did better in the treatment condition. However, it may be driven by two different mechanisms—lower achieving students may actually learn more in the treatment condition, while higher-achieving students may just be less likely to try their best on the test after working with the standard assignments. In any case, this suggests that further investigation of potential differential improvements for higher- and lower-achieving students is warranted.

One obvious limitation to Experiment 1 is the small sample size, which precluded us from accounting for the nested structure of our data through using multi-level modeling. Thus,

the main purpose of Experiment 2 was to replicate Experiment 1 with a larger sample on a wider variety of topics to examine whether students improve after using worked example assignments in real-world classrooms. As in Experiment 1, we aimed to investigate differences in improvement after use of the assignments; with the larger sample, we were able to investigate possible differences in based on both student prior knowledge and the socioeconomic (SES) level of the school, both of which are concerns for districts with highly diverse populations. Though no interaction between prior knowledge and treatment was found in Experiment 1, we still expected to find additional improvements for students with lower prior knowledge in Experiment 2, consistent with the work of Kalyuga and colleagues (2001). Investigation into potential differential improvements for schools with higher vs. lower SES levels was conducted because prior research has found the effectiveness of interventions to vary based on this factor. Some studies indicate that interventions may work less well in highly disadvantaged settings (e.g., Abdal-Haqq, 1996; Songer, Lee, & Ham, 2002); the interventions are thought to suffer due to a lack of teacher preparation or professional development (Ingersoll, 2004; Abdal-Haqq, 1996), or resources available to the teachers and students (Songer et al., 2002), both of which may hinder successful implementation of the intervention. However, in the present study, the intervention does not require extensive preparation or professional development (as the teachers conducting the experiment did not have to alter their teaching methods), nor does it require considerable resources (as all materials were provided to the teachers, and all materials are available free of charge for download on the project website, http://math.serpmedia.org/algebra_by_example/). In this context, it is plausible that the intervention could work *as well or better*, as the assignments may fill a critical gap in the

learning materials available in the classroom, and increase the ability of all students to achieve to their full potential in mathematics.

Method

Participants. 425 non-honors Algebra I students from five MSAN school districts participated. Each participating class ($N = 28$) completed one content unit (pre-algebra skills, linear equations, graphing, or quadratic equations), and individual students were randomly assigned to condition within classrooms (218 Experimental, 207 Control). Of the participating students, 30 (14 Experimental, 16 Control²) did not complete the posttest, thus their data were excluded from analysis³. The final sample consisted of 395 students (204 Experimental, 191 Control; 53% male) in 28 classrooms (16 teachers). The ethnicity breakdown was: 33% White, 40% Black, 15% Hispanic, 6% Asian, and 6% biracial. Using free or reduced lunch eligibility as a proxy, 41% of students were classified as low-SES; there was a trend toward the experimental group having a higher percentage of low-SES students (46%) than the control group (37%; $t(393) = 1.806, p = .07$). The experimental sessions were conducted in a typical course setting, with all testing done as part of normal classroom activities and all study activities administered as classroom assignments.

Measures. As in Experiment 1, the content pretests and posttests were composed of items that address the algebra content covered in the assignments and included conceptual and

² Overall attrition was 8%; differential attrition was 8% (Control) – 6% (Experimental) = 2%, which is lower than the conservative level required for maintaining minimal attrition (What Works Clearinghouse, 2014)

³ There was no difference in standardized pretest scores for students who did ($M = -.006$) and did not ($M = .08$) complete the posttest, ($t[423] = .483, p = .63$). However, Chi squared analyses revealed that a higher proportion of excluded students than expected were classified as low-SES, $X^2 (N=425) = 4.00, p = .05$. The distribution of excluded students also differed across classroom, with a higher proportion than expected coming from three of the classrooms, $X^2 (N=425) = 257.93, p < .001$. This may suggest that the teachers in those classrooms were less flexible with posttest administration and may not have allowed absent students to make it up on another day.

procedural items relevant to the content unit; samples items for all three of the new content areas may be found in Figure 3. Thus, in this experiment, there were four such tests: one on Linear Equations, one on Pre-Algebra, one on Graphing, and one on Quadratics. Internal consistency of each test was sufficient (Linear Equations: $\alpha = .83$; Pre-Algebra: $\alpha = .78$; Graphing: $\alpha = .82$; Quadratics: $\alpha = .81$); face validity for each test was also ensured, given that tests were constructed of items addressing content covered in the relevant textbook units⁴. Percentage correct scores were computed for each student at both pretest and posttest. In order to compare the effectiveness of the treatment across the units, we computed z-score transformations, separately for pretests and posttests, for each of the four units; means and standard deviations for each group on each topic unit test, plus the overall z-scores can be found in Table 2.

Instructional manipulation. The intervention was identical to that in Experiment 1 in that each student received four treatment or four control assignments during their content unit; each classroom (and therefore, each student) participated during only one content unit. The specific topics of the assignments were necessarily different for each of the content areas. The Linear Equations unit included the same four topics as in Experiment 1. The other content units were composed of the following topics: Pre-Algebra (order of operations, mathematical properties, absolute value, and fraction arithmetic), Graphing (graphing linear equations, slope, writing equations in point-slope and slope-intercept form, and function notation), and Quadratics (simplifying radicals, graphing quadratic functions, the quadratic formula, and families of

⁴ Standardized test scores were not collected for the study. However, one of the participating schools provided the students' normed percentile scores from the math portion of their most recent EXPLORE test (ACT, 2014). Individual students' normed math scores correlated positively with their scores on the Linear Equations pretest ($r[64] = .25, p = .05$), indicating concurrent validity for the Linear Equations test.

functions). All assignments were co-developed by researchers and MSAN collaborators as in Experiment 1.

Procedure. The procedure within any given content unit was identical to that of Experiment 1.

Results

As in Experiment 1, students were nested in classrooms. However, unlike Experiment 1, there were a sufficient number of clusters ($N = 28$ classrooms) to warrant the use of multi-level modeling (Maas & Hox, 2005). Thus, all analyses for Experiment 2 were conducted with hierarchical linear modeling (Raudenbush, Bryk, & Congdon, 2004). Participating classrooms were also nested in schools, and the intra-class correlations revealed that a significant amount of variance (18%) was at the school level. Unfortunately, the number of clusters was not sufficient ($N = 7$ schools) for a three-level HLM analysis (Bell, Morgan, Cromrey, & Ferron, 2010). Therefore, given the need to account for school-level variables in the most appropriate way possible, we placed a key covariate at level 2: the percentage of students in the school who are eligible for free or reduced lunch.

Thus, all analyses used student data at Level 1 and classroom data at Level 2. Level 1 included the student's z-standardized pretest and posttest scores on the content measures and condition. Level 2 included the proportion of students in the school that were eligible for free or reduced lunch, a proxy for low socioeconomic status. Descriptive statistics may be found in Table 3. With 395 participants, the minimally detectable difference at 80% probability is expected to be .142 standard deviations of change in the dependent variable for each standard deviation of change in the independent variable.

Effect of the intervention on content measures. To determine whether worked example assignments improved learning for algebra students, and whether there are differences in improvement based on either individual student's prior knowledge or schools' socioeconomic status, we conducted a series of two-level hierarchical linear models with individual students nested in classrooms. We first tested an empty model (Model 1) and determined that 23.7% of the variance in z-scored posttests was between classrooms, supporting the need for multi-level modeling. Because the intra-class correlation was 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), student prior knowledge (z-scored pretest) and school SES level (proportion of students eligible for free or reduced lunch), and the interactions between these two variables and condition. Model 2 was the intent to treat model, including only condition at Level 1. In Model 3, we included z-standardized pretest scores at Level 1 and school SES at Level 2. Finally, in Model 4, we included the interaction between z-standardized pretest scores and assignment to the treatment condition at Level 1 and a cross-level interaction between individual assignment to the treatment condition (Level 1) and school SES (Level 2).

As shown in Table 4, while Model 2 revealed no significant main effect of condition on posttest scores, Model 3 yielded a significant main effect of pretest scores on posttest scores, and Model 4 additionally yielded a significant interaction between condition and pretest scores. These results demonstrated that, while students who score better on the pretest score better on the posttest, the influence of condition varied by pretest scores such that low prior knowledge students improved more in the treatment condition (see Figure 4). The non-significant cross-level

interaction between condition and school SES demonstrated that the influence of condition did not vary based on school SES. The Akaike Information Criterion (AIC) was lower in both Model 3 (1022) and Model 4 (1023) than in Model 1 (1067) or Model 2 (1068), indicating increased fit of those models.

Discussion

The results from Experiment 2 extended the findings of Experiment 1 showing a relationship between students' prior knowledge and improved performance in the treatment condition; Low prior knowledge students showed stronger outcomes after using worked example assignments than those with higher prior knowledge. This is consistent with the work of Kalyuga and colleagues (2001), which demonstrated that low-achieving students improve more after studying worked examples than higher-achieving peers.

In Experiment 2, the main effect of condition was not significant, suggesting that worked examples were not generally beneficial for improving student performance. This may seem surprising given the main effect observed in Experiment 1, and the abundance of laboratory studies showing learning improvements after studying worked examples. Interestingly, both Experiment 1 and many of the algebra-related worked examples studies (Booth et al., 2013; Carroll, 1994; Cooper & Sweller, 1987; Sweller & Cooper, 1985) were conducted on the topic of linear equations. However, Experiment 2 included a variety of topics, not just equation-solving. It has already been established that equation-solving is an area in which students hold a high number of misconceptions and that these misconceptions are particularly detrimental to learning in this domain (Booth & Koedinger, 2008). We suspect that misconceptions may be more or less prevalent (in terms of their persistence) and problematic (in terms of their detriment to learning)

depending on the particular topic areas in Algebra; in fact, recent work suggests that this is the case (Booth, Barbieri, Eyer, & Pare-Blagoev, 2014). Thus, it is possible that our worked examples, which are particularly designed to target misconceptions, may be more or less beneficial for the general population depending on the topic area. Future work should examine this possibility with a wider range of Algebra topics.

Interestingly, the effectiveness of the assignments was also not found to vary with the schools' socioeconomic status levels. Prior research has found interventions to work less well in disadvantaged schools (Abdal-Haqq, 1996; Songer et al., 2002), likely due to reduced teacher preparation and professional development (Ingersoll, 2004; Abdal-Haqq, 1996), or available resources (Songer et al., 2002). However, in the present study, no specific teacher preparation or professional development was necessary, and all materials were provided for the schools. The fact that the intervention was equally effective for low-performing students in higher and lower-SES schools may indicate that by providing interventions that do not require schools to provide extensive professional development or resources for successful implementation may level the playing field. Regardless of the opportunities and resources typically available in a school, such interventions have the potential to impact learning, perhaps especially for low-achieving students.

Limitations

A serious limitation to the present studies is that we did not collect standardized mathematics test scores for all of the participating students. While our researcher-designed content tests were shown to be reliable and had face validity, we were only able to establish concurrent validity for one of the tests. Clearly, our results would be strengthened if validity

could be established for all of the measures. Future work should certainly aim to validate these measures. However, a recent study has shown that the impact of AlgebraByExample assignments can also be seen in students' posttest performance on released items from standardized tests (Booth, Cooper, Donovan, Huyghe, Koedinger, & Paré-Blagoev, 2015). This provides additional evidence that the AlgebraByExample assignments are increasing students' mathematical proficiency and not just performance on closely aligned measures.

Another limitation is that we cannot tell, at present, whether it is self-explaining correct examples, incorrect examples, or both which are beneficial for low-achieving students. In prior work, studying correct examples was shown to be especially beneficial for low-achieving students (Kalyuga et al., 2001), and explaining and fixing incorrect examples was shown to be less beneficial for such students (Große & Renkl, 2007). Thus, it is possible that the improvements seen in the present study were due solely to the correct examples; if this were the case, perhaps having students explain more correct examples would be more useful than having them explain a mix of correct and incorrect examples. Despite the seemingly inconsistent results with those of Große & Renkl (2007), we suspect that the incorrect examples do contribute to the low-achieving students' improvement, for several reasons. First, as mentioned previously, Große & Renkl's finding that low performing students did not improve after studying incorrect examples was tested in a context in which the students had to locate the error themselves. In the second experiment in their study, when the location of the error was highlighted, low-achieving students' performance did not suffer; in the present study, the locations of the errors were also highlighted. Second, the examples in the present study were chosen specifically to target key misconceptions students hold about the content. This could be

expected to yield stronger outcomes for low-achieving students who likely hold more of these misconceptions than higher-achieving peers. Third, previous work has shown that explaining incorrect examples may be even more beneficial to students in general than explaining correct examples (Booth et al., 2013). Further work isolating the impact of explaining correct vs. incorrect examples should be undertaken in order to test these hypotheses and ensure that the types of examples given to low-achieving students are optimal for their learning.

Educational Implications

The present study was carried out in a close collaboration between teachers, school administrators, and researchers, all of whom were aware of how intervention work often leads to null results. The team worked together to come up with ways to implement the worked examples that would not be disruptive in classrooms, would not require extensive professional development or resources, and would not ask teachers to make large changes in their teaching practices (see Booth et al.[2015] for a full account of the collaboration). Results from the present study suggest these efforts were fruitful, in that the worked example assignments were able to improve learning for some or all students, depending on the context. Future researcher-practitioner collaborations should focus on finding more innovative ways to implement interventions in classrooms; this could help ensure that students in all types of classrooms have the opportunity to learn from the intervention, and could also increase teacher and administrator buy-in when time and monetary resources are limited.

Results from the present study suggest that when used over a content unit in Algebra, example-based assignments can be beneficial in diverse populations such as urban and inner-ring suburban districts. However, it is interesting to note that while the worked example effect

translated to real-world classrooms in some contexts and for some students, the present study suggests that it does not necessarily lead to improved learning for all students in real classrooms. This is consistent with the findings of Davenport and colleagues (2007) and Dynarski and colleagues (2007), that laboratory results do not automatically translate to classrooms, and indicates that more studies with ecological validity are necessary to determine which principles translate and in which contexts. Thus, unlike what would be concluded from the abundance of laboratory studies, even well thought-out incorporation of a well-established learning principle into the complexities of the classroom environment will not necessarily lead to appreciable learning gains for all students.

Acknowledgments

Preliminary versions of this work were presented at meetings of the Eastern Psychological Association and the Cognitive Development Society. Thanks are due to the Arlington Public Schools, the Madison Metropolitan School District, the Evanston-Skokie School District, the Evanston Township School District, and the Shaker Heights School District for allowing us to collect data in their classrooms.

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

- Abdal-Haqq, I. (1996). Making time for teacher professional development. In *ERIC Digest* (#95-4). Washington, D. C.: ERIC Clearinghouse on Teaching and Teacher Education.
- Aleven, V., & Koedinger, K.R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science*, 26, 147-179.
- Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Combining fading with prompting fosters learning. *Journal of Educational Psychology*, 95, 774-783.
- Barab, S. A. & Squire, K. D. (2004). Design-based research: Putting our stake in the ground. *Journal of the Learning Sciences*, 13 (1), 1-14.
- Baroody, A. & Ginsburg, H. (1983). The effects of instructions on children's understanding of the equals sign. *The Elementary School Journal*, 84, 199-212.
- Bell, B.A., Morgan, G.B., Kromrey, J.D., & Ferron, J.M. (2010). The impact of small cluster size on multilevel models: A Monte Carlo examination of two-level models with binary and continuous predictors. *JSM Proceedings, Section on Survey Research Methods*. Vancouver, BC: American Statistical Association. \4057 –4067.
- Ben-Zeev, T., & Star, J.R. (2001). Spurious correlations in mathematical thinking. *Cognition and Instruction*, 19, 253-275.
- Booth, J.L (2011). Why can't students get the concept of math? *Perspectives on Language and Literacy*, Spring 2011, 31-35.

- Booth, J.L., Cooper, L., Donovan, M.S., Huyghe, A., Koedinger, K.R., & Paré-Blagoev, E.J. (2015). Design-based research within the constraints of practice: AlgebraByExample. *Journal of Education for Students Placed at Risk*, 20(1-2), 79-100.
- Booth, J.L., Barbieri, C., Eyer, F., & Paré-Blagoev, E.J. (2014). Persistent and pernicious misconceptions 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. <http://astro.temple.edu/~jlynn80/BoothKoedingerCogSci2008.pdf>
- Booth, J.L., Koedinger, K.R., & Siegler, R.S. (2007). [Abstract]. The effect of prior conceptual knowledge on procedural performance and learning in algebra. In D.S. McNamara & J.G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society*. Austin, TX: Cognitive Science Society.
- Booth, J.L., Lange, K.E., Koedinger, K.R., & Newton, K.J. (2013). Example problems that improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction*, 25, 24-34.
- Booth, L.R. (1984). *Algebra: Children's Strategies and Errors*. Windsor, UK: NFER-Nelson.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, 2(2), 141–178.
- Carroll, W. M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology*, 86(3), 360–367.

- Catrambone, R., & Holyoak, K.J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 1147-1156.
- Catrambone, R., & Yuasa, M. (2006). Acquisition of procedures: The effects of example elaborations and active learning exercises. *Learning and Instruction*, 16, 139-153.
- Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General*, 127(4), 355-376.
- Chang, N. (2006). Learning to discriminate and generalize through problem comparisons. Unpublished doctoral dissertation. Department of Psychology. Carnegie Mellon University.
- Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in Instructional Psychology*, Hillsdale, New Jersey: Lawrence Erlbaum Associates. 161-238.
- Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, 79(4), 347-362.
- Davenport, J. L., Klahr, D., & Koedinger, K. R. (2007). *The influence of diagrams on chemistry learning*. Paper presented at the Biennial Conference of the European Association for Research on Learning and Instruction, Hungary, Budapest.
- Department of Education. (1997). *Mathematics equals opportunity*. White Paper prepared for U.S. Secretary of Education Richard W. Riley.

- Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., Means, B., ..., & Sussex, W. (2007). *Effectiveness of Reading and Mathematics Software Products: Findings from the First Student Cohort*. Report provided to Congress by the National Center for Education Evaluation.
- Große, C.S. & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning & Instruction*, 17, 617-634.
- Hausmann, R. G. M., & VanLehn, K. (2007). Explaining self-explaining: A contrast between content and generation. In R. Luckin, K. R. Koedinger & J. Greer (Eds.), *Artificial intelligence in education: Building technology rich learning contexts that work* (Vol. 158, pp. 417-424). Amsterdam: IOS Press.
- Hilbert, T. S., Renkl, A., Kessler, S. & Reiss, K. (2008). Learning to prove in geometry: Learning from heuristic examples and how it can be supported. *Learning and Instruction*, 18, 54-65.
- Ingersoll, R. M. (2004). *Why do high-poverty schools have difficulty staffing their classrooms with qualified teachers? (Report prepared for Renewing Our Schools, Securing Our Future – A National Task Force on Public Education)*. Washington, DC: The Center for American Progress and the Institute for America's Future. Retrieved June 1, 2013, from <http://www.americanprogress.org/site/pp.asp?c=biJRJ8OVF&b=252682>.
- Kalyuga, S., Chandler, P., & Sweller, J. (2001a). Learner experience and efficiency of instructional guidance. *Educational Psychology*, 21(1), 5-23.
- Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001b). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, 93(3), 579–588.

- Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics*, 12, 317-326.
- Kim, R., Weitz, R., Heffernan, N. & Krach, N. (2009). Tutored Problem Solving vs. “Pure”: Worked Examples In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 3121-3126). Austin, TX: Cognitive Science Society.
- 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.
- Koedinger, K.R., Booth, J.L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science*, 342, 935-937
- Küchemann, D. (1978). Children’s understanding of numerical variables. *Mathematics in School*, 7, 23-26.
- Kulik, J.A. (1994). Meta-analytic studies of findings on computer-based instruction. In E.L. Baker and H.F. O’Neil, Jr. (Eds), *Technology Assessment in Education and Training* (pp. 9-33). Hillsdale, NJ: Lawrence Erlbaum.
- Ladson-Billings, G.J. (1995). Toward a theory of culturally relevant pedagogy. *American Education Research Journal*, 35, 465-491.
- Maas, C.J.M. & Hox, J.J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1(3). 86–92.
- Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology*, 81, 240-246.

- Mayer, R. (2005). Cognitive theory of multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). Cambridge: Cambridge University Press.
- McLaren, B.M., Lim, S., Gagnon, F., Yaron, D., and Koedinger, K.R. (2006). Studying the effects of personalized language and worked examples in the context of a web-based intelligent tutor. In *The Proceedings of the 8th International Conference on Intelligent Tutoring Systems* (pp. 318-328). New York: Springer.
- Moreno, R., Reisslein, M., & Ozogul, G. (2009). Optimizing worked-example instruction in electrical engineering: The role of fading and feedback during problem-solving practice. *Journal of Engineering Education*, 98, 83-92.
- Murphy, R., Penuel, W., Means, B., Korbak, C., & Whaley, A. (2001). *E-DESK: A Review of Recent Evidence on the Effectiveness of Discrete Educational Software*. Menlo Park, CA: SRI International.
- National Mathematics Advisory Panel. (2008). *Foundations for Success: The Final Report of the National Mathematics Advisory Panel*, U.S. Department of Education: Washington, DC.
- Ohlsson, S. (1996). Learning from error and the design of task environments. *International Journal of Educational Research*, 25(5), 419-448.
- 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.
- Paas, F., & Van Merriënboer, J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, 86(1), 122-133.

- Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., and 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.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods* (2nd). Thousand Oaks, CA: Sage Publications.
- Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2004). HLM (Version 6)[Computer software]. Lincolnwood, IL: Scientific Software International.
- Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science: A Multidisciplinary Journal*, 21(1), 1-29.
- Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and Instruction*, 12(5), 529-556.
- Renkl, A., Atkinson, R., Maier, U., & Staley, R. (2002). From example study to problem solving: Smooth transitions help learning. *Journal of Experimental Education*, 70(4), 293-315.
- Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, 77, 1-29.
- Ross, B.H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 456-468.
- 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: Cambridge University Press.

- Schorr, T., Gerjets, P., Scheiter, K., & Laouris, Y. (2002). Designing sets of instruction examples to accomplish different goals of instruction. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 810--815). Mahwah, NJ: Erlbaum.
- Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., & Salden, R. J. C. M. (2009). The worked-example effect: Not an artifact of lousy control conditions. *Computers in Human Behavior*, 25, 258-266.
- 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: Cambridge University Press.
- Siegler, R. S., & Chen, Z. (2008). Differentiation and integration: Guiding principles for analyzing cognitive change. *Developmental Science*, 11, 433-448.
- Songer, N. B., Lee H., & Kam, R. (2002). Technology-rich inquiry science in urban classrooms: What are the barriers to inquiry pedagogy? *Journal of Research in Science Teaching*, 39(2), 128-150.
- Steele, C. M. (1992). Race and the schooling of Black Americans. *Atlantic Monthly*, April, 68--78.
- 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.
- 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.

- Tarmizi, R.A. and Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology*, 80 (4) 424-436
- Vlassis, J. (2004). Making sense of the minus sign or becoming flexible in ‘negativity’. *Learning and Instruction*, 14, 469-484.
- Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, 7(1), 1-39.
- What Works Clearinghouse (2014). *Procedures and Standards Handbook, Version 3.0*. Retrieved from <http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19>.
- Zhu, X., & Simon, H.A. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction*, 4, 137-166.

Table 1.
Experiment 1: Means and Standard Deviations for Content Scores at Pretest and Posttest by Condition and Prior Knowledge (High vs. Low)

| | Pretest | Posttest | Analysis |
|----------------------|-----------|-----------|--------------------------|
| Treatment Condition | .68 (.14) | .72 (.12) | $t(25) = -1.53, ns$ |
| High Prior Knowledge | .79 (.06) | .77 (.14) | $t(11) = .52, ns$ |
| Low Prior Knowledge | .58 (.11) | .68 (.08) | $t(13) = -2.39, p < .05$ |
| Control Condition | .69 (.12) | .64 (.16) | $t(24) = 1.68, ns$ |
| High Prior Knowledge | .79 (.05) | .70 (.16) | $t(12) = 2.70, p < .05$ |
| Low Prior Knowledge | .58 (.06) | .59 (.14) | $t(11) = -1.45, ns$ |

Note. Mean(SD)

Table 2.

Experiment 2: Means and Standard Deviations for Content Scores by Condition and Topic, and Z-standardized Content Scores Across Topic

| | Pretest | Posttest | Analysis |
|-----------------------|-------------|-------------|----------------------------|
| Pre-Algebra | | | |
| Treatment | .48 (.17) | .53 (.15) | $t(24) = -1.70, p = .10$ |
| Control | .48 (.15) | .60 (.14) | $t(20) = -4.72, p < .001$ |
| Linear Equations | | | |
| Treatment | .64 (.16) | .68 (.13) | $t(84) = -2.45, p < .05$ |
| Control | .64 (.17) | .66 (.15) | $t(78) = -1.32, ns$ |
| Graphing | | | |
| Treatment | .41 (.11) | .65 (.12) | $t(59) = -12.96, p < .001$ |
| Control | .41 (.11) | .61 (.15) | $t(54) = -8.81, p < .001$ |
| Quadratics | | | |
| Treatment | .44 (.11) | .77 (.17) | $t(33) = -9.72, p < .001$ |
| Control | .42 (.12) | .78 (.23) | $t(35) = -10.68, p < .001$ |
| Overall Z-scores | | | |
| Treatment | .02 (.98) | .04 (.92) | $t(203) = -.31, ns$ |
| Control | -.02 (1.01) | -.04 (1.07) | $t(190) = .32, ns$ |
| <i>Note.</i> Mean(SD) | | | |

Table 3.

Descriptive statistics for Level 1 and Level 2 variables in Experiment 2

| | N | Mean | Standard Deviation | Minimum | Maximum |
|---------------------------------------------------------------------|-----|------|-----------------------|---------|---------|
| <i>Student Level Variables: Level 1</i> | | | | | |
| Treatment | 395 | 0.52 | 0.50 | 0.00 | 1.00 |
| Pretest Score (Z-standardized) | 395 | 0.00 | 1.00 | -3.93 | 2.90 |
| Posttest Score (Z-standardized) | 395 | 0.00 | 1.00 | -4.70 | 2.31 |
| <i>Classroom Level Variables: Level 2</i> | | | | | |
| School SES (proportion of students eligible for free/reduced lunch) | 28 | 0.39 | 0.08 | 0.29 | 0.63 |

Table 4.
Predictors of Posttest Performance in Experiment 2

| | Model 1 (Unconditional model) | Model 2 (Intent to treat model) | Model 3 (Main effects model) | Model 4 (Interaction model) |
|------------------------------------------------------------------------------------------------------------------------------------|----------------------------------|------------------------------------|---------------------------------|--------------------------------|
| Fixed Effects | | | | |
| <i>Student-Level</i> | | | | |
| Treatment | -- | .10 (.08) | .08 (.09) | .08 (.09) |
| <i>Classroom-Level</i> | | | | |
| Intercept | -.08 (.11) | -.13 (.12) | -.11 (.10) | -.11 (.10) |
| School SES (proportion eligible for free/reduced lunch) ^a | -- | -- | 1.44 (1.00) | 1.40 (.92) |
| School SES ^a x Treatment interaction | -- | -- | -- | .06 (1.25) |
| Random Effects | | | | |
| <i>Student-Level</i> | | | | |
| Pretest Score (Z-Standardized) | -- | -- | .33** (.06) | .41** (.08) |
| Pretest x Treatment Interaction | -- | -- | -- | -.16* (.07) |
| <i>Variance Components</i> | | | | |
| Classroom-Level | .239 | .241 | .146 | .149 |
| Student-Level | .771 | .768 | .692 | .685 |
| Proportion Reduction in Level-1 Variance (from Model 1) | -- | .004 | .102 | .112 |
| <i>Notes.</i> β (SE); ^a Predictor is grand-mean centered; * $p < .05$; ** $p < .01$; --not included in model | | | | |

Figure 1: Excerpts from the treatment and control versions of the assignment on Solving Multi-Step Equations for Experiment 1

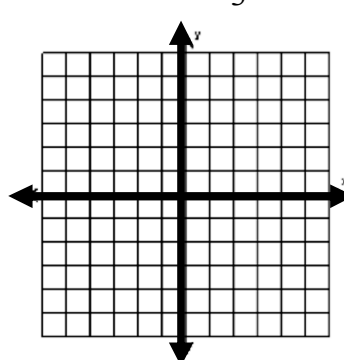
| | | |
|----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------|
| Worked Example (Treatment) | <p>3. $\frac{3}{4x-6} = 5$</p> <p><u>Correct Example</u></p> <p>Sasha solved this problem correctly. Here are the steps she used to solve the problem:</p> $\frac{3}{4x-6} = 5 \cdot \frac{4x-6}{4x-6} \quad \left. \vphantom{\frac{3}{4x-6}} \right\} *$ $3 = 5(4x-6)$ $3 = 20x - 30$ $\frac{33}{20} = \frac{20x}{20}$ $\frac{33}{20} = x$ <p>Why did Sasha multiply both sides by $(4x-6)$ in the highlighted step(*)?</p> | <p>4. $\frac{3}{6-4x} = 5$</p> |
| | <p>5. $3x = 4x - 6 + 5$</p> <p><u>Incorrect Example</u></p> <p>Umi tried to solve this problem, but she didn't do it correctly. Here is her first step to solve the problem:</p> $3x = 4x - 6 + 5$ $3x = 4x - 11$ <p>What was Umi's first step to solve the problem?</p> <p>Why didn't Umi's first step keep the equation in balance?</p> | <p>6. $3 = 4 + 6x - 5x$</p> |

| | | |
|---------|-------------------------|-------------------------|
| Control | 3. $\frac{5}{2d-4} = 9$ | 4. $\frac{4}{8-7r} = 5$ |
| | 5. $2k = 5k - 9 + 6$ | 6. $8 = 7 + 4v - 3v$ |

Figure 2: Sample content assessment items for Experiment 1

| Conceptual | Procedural | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|----------|-----|----|----|----------|-----|----|----|----------|-----|----|----|-----------|-----|----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|-----|----|----|-----|-----|----|----|----------|-----|----|----|-------|-----|----|-------------------------------------------------------|
| <p>State whether each of the following is a like term for $8c$:</p> <table><tr><td>a.</td><td>-5</td><td>Yes</td><td>No</td></tr><tr><td>b.</td><td>$2c^2$</td><td>Yes</td><td>No</td></tr><tr><td>c.</td><td>$6c$</td><td>Yes</td><td>No</td></tr><tr><td>d.</td><td>$3d$</td><td>Yes</td><td>No</td></tr><tr><td>e.</td><td>$\frac{c}{3}$</td><td>Yes</td><td>No</td></tr><tr><td>f.</td><td>8</td><td>Yes</td><td>No</td></tr><tr><td>g.</td><td>$5(c+1)$</td><td>Yes</td><td>No</td></tr><tr><td>h.</td><td>$-4c$</td><td>Yes</td><td>No</td></tr></table> | a. | -5 | Yes | No | b. | $2c^2$ | Yes | No | c. | $6c$ | Yes | No | d. | $3d$ | Yes | No | e. | $\frac{c}{3}$ | Yes | No | f. | 8 | Yes | No | g. | $5(c+1)$ | Yes | No | h. | $-4c$ | Yes | No | <p>Solve for x: $2 = \frac{8}{3 - 4x}$</p> |
| a. | -5 | Yes | No | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| b. | $2c^2$ | Yes | No | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| c. | $6c$ | Yes | No | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| d. | $3d$ | Yes | No | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| e. | $\frac{c}{3}$ | Yes | No | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| f. | 8 | Yes | No | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| g. | $5(c+1)$ | Yes | No | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| h. | $-4c$ | Yes | No | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Grandpa's age is 2 less than 6 times Miguel's age, m. State whether each of the following expressions represents Grandpa's age:</p> <table><tr><td>a.</td><td>$6m + 2$</td><td>Yes</td><td>No</td></tr><tr><td>b.</td><td>$2 - 6m$</td><td>Yes</td><td>No</td></tr><tr><td>c.</td><td>$6m - 2$</td><td>Yes</td><td>No</td></tr><tr><td>d.</td><td>$-2 + 6m$</td><td>Yes</td><td>No</td></tr></table> | a. | $6m + 2$ | Yes | No | b. | $2 - 6m$ | Yes | No | c. | $6m - 2$ | Yes | No | d. | $-2 + 6m$ | Yes | No | <p>Elpheria's monthly cell phone bill is \$25.00, plus an additional 8 cents (\$.08) for every text message she sends. Write an equation that you could use to help you find out how many texts she sent this month if her bill was \$34.60. You <i>do not</i> need to solve the equation. Remember to define your variables.</p> | | | | | | | | | | | | | | | | |
| a. | $6m + 2$ | Yes | No | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| b. | $2 - 6m$ | Yes | No | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| c. | $6m - 2$ | Yes | No | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| d. | $-2 + 6m$ | Yes | No | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Figure 3: Sample content assessment items for Experiment 2

| | Conceptual | Procedural |
|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Pre-Algebra | <p>State whether each of the following is equivalent to $x + 4 - 2 + x$:</p> <p>a. $(x + 4) - (2 + x)$ Yes No</p> <p>b. $4 + x - 2 + x$ Yes No</p> <p>c. $x + (4 - 2) + x$ Yes No</p> <p>d. $x + 4 - x + 2$ Yes No</p> <p>e. $(x + 4) + (-2 + x)$ Yes No</p> <p>f. $x + 4 + x - 2$ Yes No</p> <p>g. $x + 2(2 - 1) + x$ Yes No</p> | <p>Find the quotient for the expression and write in simplest form. Show all of your work: $\frac{4}{5} \div -2$</p> |
| Graphing | <p>State whether each of the following is true for the line: $(y - 3) = 2(x + 1)$</p> <p>a The line goes through (1,-3) Ye N s o</p> <p>b The slope-intercept form of the line is $y = 2x + 5$ Ye N s o</p> <p>c The line has a slope of 2 Ye N s o</p> <p>d The line goes through (3, -1) Ye N s o</p> <p>e The line has a slope of $\frac{1}{2}$ Ye N s o</p> <p>f The line goes through (-1, 3) Ye N s o</p> <p>g The slope-intercept form of the line is $y = 2x - 1$ Ye N s o</p> | <p>Find the x- and y-intercepts. Then use them to graph the equation: $y = 2 - \frac{1}{3}x$</p>  |

| | | | | | | | | | | | | | | | | |
|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|-------------|-------------|----------------------------------------------------------------------------------|-----|----|-----|----|----|--------|-----------|-------------|-----------|-------------|
| Quadratics | Identify the type of each function: | | | | | Solve for z using the quadratic formula. Show all of your work. $z^2 - 4z - 3$ | | | | | | | | | | |
| | a. | $y = x^2 + 6$ | Linear | Quadratic | Exponential | | | | | | | | | | | |
| | b. | <table border="1"><tr><td>x</td><td>2</td><td>4</td><td>6</td><td>8</td></tr><tr><td>y</td><td>1</td><td>4</td><td>7</td><td>10</td></tr></table> | x | 2 | 4 | | 6 | 8 | y | 1 | 4 | 7 | 10 | Linear | Quadratic | Exponential |
| | x | 2 | 4 | 6 | 8 | | | | | | | | | | | |
| | y | 1 | 4 | 7 | 10 | | | | | | | | | | | |
| | c. | $y = 2^x + 6$ | Linear | Quadratic | Exponential | | | | | | | | | | | |
| e. | <table border="1"><tr><td>x</td><td>3</td><td>4</td><td>5</td><td>6</td></tr><tr><td>y</td><td>11</td><td>18</td><td>27</td><td>38</td></tr></table> | x | 3 | 4 | 5 | 6 | y | 11 | 18 | 27 | 38 | Linear | Quadratic | Exponential | | |
| x | 3 | 4 | 5 | 6 | | | | | | | | | | | | |
| y | 11 | 18 | 27 | 38 | | | | | | | | | | | | |
| f. | $y = 2x + 6$ | Linear | Quadratic | Exponential | | | | | | | | | | | | |

Figure 4: Interaction between treatment and prior knowledge on posttest performance in Experiment 2

