

# Evoking learning by examples through reducing misconceptions and highlighting procedures

Authors: Christina Areizaga Barbieri<sup>1</sup> and Elena Silla<sup>1</sup>

1. University of Delaware, 19716

Running Head: MECHANISMS OF LEARNING FROM EXAMPLES

*Keywords:* algebra learning, incorrect examples, misconceptions, procedural knowledge, worked examples

*Author note:* The authors thank Courtney Martin, Emily O’Hara, and Camille Roth for assistance in data entry and coding as well as participating school districts.

First Submission: June 2, 2022

First Decision: March 7, 2023

Revision Submitted: June 5, 2023

*Corresponding author:*

Christina Areizaga Barbieri, Ph.D.

barbieri@udel.edu

University of Delaware

College of Education and Human Development

113 Willard Hall Education Building

Newark, DE 19716

**Abstract**

Prior research highlights a positive effect of incorrect worked examples on mathematics learning. Yet the mechanisms underlying these benefits are unclear. To investigate potential mechanisms of the benefits of various worked example types, we examined process data from a previously published classroom-based experiment (Barbieri & Booth, 2020). More specifically, we analyzed students' explanations made while explaining worked examples in three varying example conditions as well as students' problem-solving errors made when solving problems. These data operationalize two potential mechanisms: a reduction of misconceptions (i.e., fewer targeted conceptual errors), and an increase in principled algebra knowledge (i.e., explanations focusing on principles underlying procedures). Mediation analyses revealed both as important mechanisms of varying effects. A reduction of misconceptions explained greater benefits of all three worked example conditions, compared to a problem-solving control, on an algebra concepts posttest. More principled explanations of procedures explained the benefits of incorrect worked examples on problem-solving at posttest compared to the two other example conditions. These findings help explain differential findings in prior work by example type and may elucidate potential avenues for errorful instruction.

*Errorful* learning techniques—or techniques that capitalize on errors to promote learning—can be effective for learning (Metcalfe, 2017 for a review). Many techniques fall under this umbrella term of errorful learning, including those that provoke uncertainty, those that ask students to confront discrepant ideas, and those that ask students to study and/or correct mathematical errors. Errorful learning techniques sometimes rely on students learning from their own errors or, in the case of incorrect worked examples, learning from a fictitious students' errors. This specific form of errorful learning is especially effective for promoting algebra learning (Barbieri & Booth, 2016, 2020; Booth et al., 2015a; Booth et. al., 2015b). Though we know that errorful learning improves algebra performance, mechanisms underlying these benefits are unclear. There is a critical need to understand *how* students learn from instructional experiences that yield improved performance in algebra. Without this knowledge, recommendations on what instructional techniques work for students with diverse learning needs would be incomplete, incorrect, or even detrimental to student learning. Thus, understanding the mechanisms underlying errorful learning is crucial not only for understanding student learning but for designing effective future interventions and instructional tools to help all learners.

In the current study, we analyze data collected as part of a prior study (Barbieri & Booth, 2020) to determine mechanisms underlying errorful learning. These previous findings showed *that* studying and explaining incorrect examples improved middle school algebra students' equation-solving skills to a greater extent than a problem-solving control condition (Barbieri & Booth, 2020) but did not reveal *why* or *how*. Incorrect worked examples were as effective as two other forms of correct examples at improving algebraic feature knowledge and all three example-types outperformed a problem-solving control on equation solving, which is consistent with previous studies on errorful learning. We expand on this previous finding by examining two

potential mechanisms underlying these potential benefits for equation-solving and algebraic feature knowledge. By examining students' errors as well as their written explanations, we hope to better explain why learning from errors is beneficial for many students and compare these mechanisms to those underlying examples focused on correct information.

### ***Incorrect Worked Examples***

An accumulating body of evidence demonstrates a generally positive effect of studying errors in the form of incorrect worked examples, or worked out problem-solutions that display a common mistake, on math learning when these examples are studied alone or in combination with correct examples (e.g., Adams et al., 2014; Barbieri & Booth, 2016; Booth et al., 2015b; Durkin & Rittle-Johnson, 2012). Students are asked to study, explain, and sometimes correct the error displayed before completing similar practice problems on their own. In one study, students who studied incorrect examples of place value on a computerized tutor showed longer-lasting learning effects than students who only solved problems (Adams et al., 2014). Similarly, Barbieri and Booth (2016) found that students benefitted from working with both incorrect and correct worked examples, but that incorrect worked examples were particularly helpful for students with low prior knowledge. Other studies have shown that a combination of correct and incorrect worked examples reduces mathematical misconceptions (e.g., Barbieri et al., 2019; Booth et al., 2015b; Durkin & Rittle-Johnson et al., 2012). Studying a combination of correct and incorrect worked examples has been shown to increase students' conceptual understanding of algebra (Booth et al., 2015b) and conceptual and procedural understanding of decimal magnitudes (Durkin & Rittle-Johnson, 2012).

Although the above findings show that learning from incorrect worked examples is helpful for math learning, the reason underlying this benefit is unknown. There have been some

proposed mechanisms explaining this benefit yet no published study has assessed these mechanisms empirically. For example, Barbieri and Booth (2016) found that incorrect worked examples were particularly beneficial for students with low prior knowledge. They posited that one reason for this finding was incorrect worked examples highlight common errors and draw attention to the parts of the problems that might be incorrect, thus supporting students with low prior knowledge by making those errors salient (Barbieri & Booth, 2016). Additionally, Durkin and Rittle-Johnson (2012) discussed the possibility that comparing correct and incorrect worked examples helped students distinguish between incorrect and correct problem-solving procedures, thus enhancing both conceptual and procedural knowledge. They also suggested that increased discussions of correct concepts, seen through students' written explanations, may have led to increased conceptual knowledge. However, this proposed explanation was not statistically assessed as a mechanism (i.e., mediator). An understanding of how these types of materials yield benefits would not only be theoretically meaningful but would also support the proper development of errorful and example-based learning materials that optimize their effects for all learners.

### ***Overview of Overarching Study (Barbieri & Booth, 2020)***

The current study focuses on exploring the potential mechanisms behind the overall benefits found in our prior study (Barbieri & Booth, 2020). In this prior study, studying and explaining incorrect examples improved middle school algebra students' equation-solving skills more than a problem-solving control. Three different example types (e.g., *Correct Explanation*, *Displayed Error*, *Potential Error*) that varied by correctness and explanation type equally improved their algebraic feature knowledge, compared to a problem-solving control. We provide

a brief overview of the condition differences here. Further details of the study are found in the methods section and in Barbieri and Booth (2020).

In the *Correct Explanation* condition students studied and explained correct concepts and procedures displayed within correct worked examples. Two additional worked examples conditions relied on errorful learning strategies – focusing on common computational and conceptually-based errors – in distinct ways. In the *Displayed Error* condition students studied and explained errors displayed in incorrect worked examples. In the *Potential Error* condition students studied correct worked examples and then answered explanation prompts that primed them to consider a specific potential error that could occur in this problem-type. Finally, students in the *Problem-solving Control group* completed the same problems presented in the example conditions but without examples and explanation prompts.

In the original study design, we included a *Potential Errors* condition for several reasons. First, because practitioners can be hesitant to display errors in the classroom but prior work emphasizes the effectiveness of studying incorrect examples, we attempted to design a learning experience that relied on errorful learning ideas without directly displaying the errors themselves. Further, practitioners may be concerned that students need to be provided with correct information to refer to when attempting to explain an error. Prior work has supported this with side-by-side comparisons of correct and incorrect examples (Durkin et al., 2012). However, some work highlights that this comparison design is difficult for students with low prior knowledge (Rittle-Johnson et al., 2009). Thus, this new *Potential Error* condition design was an attempt to address the possible need of providing learners correct information to reference when explaining an error. That is, we attempted to evoke the same benefits of errorful learning gained from incorrect examples by just prompting students to consider the commonly made potential

error. Our findings from Barbieri and Booth (2020) demonstrated that although the Potential Error condition was significantly more effective than the Problem-solving Control at improving algebraic feature knowledge in particular, this improvement was equivalent to those yielded by the Displayed Error condition and the Correct explanation condition. That is, we yielded a general worked examples effect regardless of the correctness or errors reflected upon in these three examples for the more conceptually-based outcome measure (i.e., algebraic feature knowledge). However, only the Displayed Errors condition was more effective than the problem-solving control at improving equation-solving. A comprehensive reporting of the overarching study and findings can be found in Barbieri and Booth (2020). This prior publication focuses on quantifying the effects of each condition but does not explore potential *mechanisms* that may explain such differential effects. The current study is a fine-grained analysis of the process data (i.e., student work on the materials) collected during the initial study that will be used to elucidate potential mechanisms of the effects of learning from errors. Understanding *how* a particular instructional technique or tool works allows for proper design of additional tools based on that technique or method. That is, understanding how something works ensures that when extending its use in new ways or in new settings, we do not inadvertently remove a key component of its design necessary for its proper function or use the tool for aims that are not supported by the mechanism(s). For example, let us assume we find that incorrect worked examples improve equation-solving skills through an increased focus on procedures as evidenced by the primary focus of student explanations. We might then encourage practitioners to use these types of examples with explanation prompts to improve students' equation-solving skills. Additionally, we would not need to prompt practitioners to emphasize or scaffold students' conceptually-based explanations with these materials and rather explain that procedurally-based

explanations are not only acceptable but expected to yield the intended effects. Alternatively, if we lacked the knowledge that this example-type evokes more procedurally-based explanations that lead to improvements in equation-solving, we (or practitioners) may be unnecessarily concerned with students who provide such explanations rather than conceptual explanations and try to intervene on those explanation-types. This could then potentially attenuate or even eradicate the positive effects of this particular example-type on equation-solving skills.

### ***Potential Mechanisms of Varying eEffects***

We expect that the varying effects found in Barbieri and Booth (2020) can be explained by varying learning processes that occur while working with the varying experimental materials. We analyze the type and frequency of errors students made on problems paired with examples or within the problem-solving condition as well as explanations provided while working with study materials. We use these data to operationalize potential mechanisms of the varying effects found on feature knowledge and equation-solving skills in Barbieri and Booth (2020). In this analysis, we consider two potential mechanisms: 1) a reduction of misconceptions and 2) a difference in the type of knowledge evoked by example-type (i.e., conceptually- vs. procedurally-focused). These mechanisms are further detailed below.

#### ***Reduction of Misconceptions***

Misconceptions can be described as strongly-held beliefs about the target content that are incorrect or only partially correct. Sometimes these misconceptions are due to overgeneralization of knowledge from the domain to new content (e.g., McNeil, 2014). Other times they are simply based on naïve ideas about the content that are unrelated to correct information (e.g., van Loon et al., 2015). Mathematical misconceptions in particular are deeply rooted and can persist or even return after being addressed (e.g., Cook et al., 2008; McNeil & Alibali, 2000) which can impact

what students gain from instruction (Kendeou & van den Broek, 2005; Resnick, 1983).

Misconceptions can interfere with knowledge acquisition and refinement if not appropriately addressed in instruction (McNeil, 2014).

Many methods proposed to address misconceptions rely on provoking uncertainty (e.g., Overoye & Storm, 2015) and encouraging the learner to resolve the uncertainty. This is a useful approach for reducing misconceptions that are pervasive in math (D'Mello & Graesser, 2014; Gorsky & Finegold, 1994). Errorful learning techniques such as studying and explaining errors within worked examples might make misconceptions more salient. Prior work has found that learning from incorrect worked examples either alone (Barbieri & Booth, 2016) or in conjunction with correct examples (Barbieri & Booth, 2020; Booth al., 2015b) is associated with enhanced benefits, particularly for students with low prior knowledge. Thus, one potential explanation for the added benefit seen from errorful learning experiences, especially for students with low prior knowledge, is that carefully designed errorful learning experiences may address more salient or persistent misconceptions held. In the current study, we examine whether the increase in algebraic feature knowledge seen in students in the three worked example conditions, but not in the problem-solving control, can be explained by a reduction in misconceptions, operationalized by making fewer targeted errors when solving problems.

#### *Evoking Different Types of Knowledge: Procedural or Conceptual*

A potential explanation for the particular benefits of the *Displayed Error* condition on equation-solving in particular is an increased attention to procedures. Algebraic equation-solving requires both procedural knowledge, or steps and strategies on how to solve the problem, and conceptual knowledge, or the reasons underlying the procedures (Crooks & Alibali, 2014). Procedural and conceptual knowledge are intertwined, with procedural knowledge predicting

conceptual knowledge and vice versa (Rittle-Johnson et al., 2015). An increase in either procedural or conceptual understanding may be predictive of equation-solving yet focusing students' attention on procedures through intentional design of examples and their corresponding explanation prompts may provide a more direct transfer to procedurally-focused skills such as equation-solving.

Studying worked examples is an effective way to increase procedural knowledge, as worked examples reduce procedural demands on working memory (Sweller, 1998), allowing students to focus on learning the procedures. Studying incorrect worked examples and explaining why the error is incorrect might increase students' understanding of the problem-solving procedures. Siegler's (1996) *overlapping waves theory* maintains that individuals know and use a variety of strategies for solving problems, and those strategies compete for use each time a student encounters a problem. Studying problem-solving or strategy errors can help learners accept that certain strategies are wrong and prompt them to construct and strengthen other correct strategies. Heemsoth and Heinze (2014) support this idea by revealing increases in sixth graders' *negative knowledge* of fractions when working with incorrect worked examples alone compared to correct examples alone. *Negative knowledge*, or knowledge of incorrect strategies and (i.e., what *doesn't* work) is important for general problem-solving (Gartmeier et al., 2008). Negative knowledge improves problem-solving success by increasing one's certainty in the correct procedure while allowing one to avoid incorrect procedures. In the current study, we examine whether the increase in equation-solving skill seen in the *Displayed Error* condition in particular, but *not* the other two worked example conditions, can be explained by an increased focus on procedures, operationalized by an increased emphasis on procedures in students' written explanations.

## Overview of the Current Study

This current study provides a deeper analysis of the potential mechanisms underlying the benefits found in our previous work (Barbieri & Booth, 2020). Specifically, the current study delves into the previously unexplored process data of the type and quantity of errors students made while working with the different materials as well as the differences in the types of explanations students provided in response to the explanation prompts.

The current study aims to test two mechanisms. First, we test whether a reduction in misconceptions explains the increase in algebraic feature knowledge associated with the three worked example conditions (Aim 1). In the prior study we found that the three worked examples conditions outperformed the *Problem-solving Control* on the algebraic feature knowledge measure. We propose that the three worked examples conditions may have yielded a reduction in misconception prevalence compared to the *Problem-solving Control* that led to the marked improvements in feature knowledge. We then operationalized this mechanism as the proportion of targeted errors made by each student on the practice problem paired with each example. We tested this measure as a mediator of the effects of condition to address our first aim.

Next, we test whether an increased focus on procedures explains the performance of students in the *Displayed Error* condition in particular on the equation-solving posttest (Aim 2). In the prior study, we found that the *Displayed Error* condition in particular outperformed the *Problem-solving control* on the equation-solving measure. The *Correct Explanation* condition and the *Potential Error* condition did not differ overall in their equation-solving performance at posttest from either the displayed error condition or the control group. We propose that an increased focus on problem-solving procedures evoked by the *Displayed Error* condition is what explains this condition's particular benefit on equation-solving skills at posttest. We

operationalized this mechanism as the proportion of procedural and conceptual explanations students provided when explaining worked examples. We then added these two measures into the model predicting post-equation-solving scores and tested them as mediators of the effects of conditions to address our second aim. Procedures for the overarching study are briefly summarized below before detailing the specifics of the current analysis.

## **Methods**

### ***Participants***

The current study was conducted in a racially and ethnically diverse public middle school located in the U.S. Midwest. Students ( $N = 206$ ; 104 female, 102 male) were from eleven Algebra I classrooms taught by three different teachers. The overall sample included White (62.8%), Black (23.2%), Asian (6.8%), multi-racial (5.8%), and Hispanic (1.4%) students, as categorized by their school district records. Students receiving free or reduced lunch, a common proxy of socioeconomic status, made up 14.1% of the sample. Although specific grade level and age were not provided for individual students due to privacy purposes, most students in the study were in the eighth grade. Eighth graders in the U.S. are typically between 13 and 14 years old. Students were randomly assigned within classroom to one of four conditions: *Problem-solving Control* ( $n = 50$ ), *Correct Explanation* condition ( $n = 51$ ), *Potential Error* condition ( $n = 53$ ), and *Displayed Error* condition ( $n = 52$ ). The current analysis focuses primarily on *process data*, which is student problem-solving and explanation data collected through the assignments that they completed during the experimental manipulation representing their learning process (as opposed to pre- or posttest data). Paper data was received by the project team via mail in separate shipments (i.e., pretests first, assignments next, posttests last). Participating teachers sent 206 students' assessment data but sent only 192 assignment packets. Thus, we do not have process

data for 7% of the original sample. However, all participants are retained in our analytical models as we use full information maximum likelihood (FIML) estimation that allows us to estimate models using all information available to us and without dropping students. This approach is described further below.

### ***Materials***

Pre- and posttests assessed students' algebraic feature knowledge and their ability to solve quadratic equations using the four methods demonstrated in the assignments (i.e., the quadratic formula, factoring, square root, graphing).

#### *Algebraic Feature Knowledge*

The measure of algebraic feature knowledge included 23 items that demonstrated a students' understanding of important features in a problem, such as coefficients (samples in Figure 1). All problems on this measure required conceptual understanding of algebraic features within quadratic equations. However, they were in a different format than the problems in the assignments. For example, one item asks, "*State whether each of the following is true for the quadratic function  $y = -x^2 - 2x + 3$ .*" Students needed to determine whether the vertex would be a maximum on the graph of this quadratic function, which would require the student to understand the meaning of the coefficient of  $x^2$ . Internal consistency at pre- and post-test was sufficient ( $\alpha = .696$ ;  $\alpha = .820$ ).

#### *Equation-solving Skill*

The measure of equation-solving included 16 items that demonstrated students' ability to solve quadratic equations using the methods demonstrated in the assignments (Samples in Figure 1). Some questions were similar in structure and format to problems displayed in the examples. For example, one question asks students to, "*Solve the quadratic equation by factoring. Show all*

of your work.  $0 = x^2 - 6x - 27$ .” Other problems relied on equation-solving skills targeted in the assignments (e.g., solving using the quadratic formula, solving by factoring, etc.) but were presented in a different format. For example, one question asks students, “A ball falls off of a ledge. The equation  $0 = -2t^2 + 10t + 0$  gives the time  $t$  in seconds when the ball hits the ground (height of 0 feet). Use the quadratic equation to find out how many seconds it took for the ball to hit the ground.” Internal consistency at pre- and posttest was sufficient ( $\alpha = .766$ ;  $\alpha = .820$ ).

### *Worked Example Assignments*

The experimental manipulation was within the four assignments that students completed, each designed to be completed in about one class period (~45 min). The four assignments focused on solving quadratic equations by using (1) the quadratic formula, (2) factoring, (3) square root, and (4) graphing.

To test the effect of studying errors on learning, there were two worked examples conditions that promoted error study and/or explanation (*Displayed Error* and *Potential Error*) and two conditions that did not, one of which was a problem-solving condition and another that was a traditional worked examples condition (*Correct Explanation*). Each worked example across the three example conditions was paired with an isomorphic problem (i.e., “Your Turn” problems) to be completed after studying the example and completing the explanation prompts. Each example targeted at least one specific error (i.e., *Displayed Error*, *Potential Error*) or the corresponding correct concept (i.e., *Correct Explanation*). Examples of targeted errors are in Table 1. Each worked example assignment included three example-problem pairs so each student studied a total of 12 worked examples and completed a total of 12 practice problems. The *Problem-solving Control* group completed 12 practice problem pairs (24 problems total) identical to the problems in the worked examples conditions but without the worked examples or

self-explanation prompts. Sample problem pairs or example-problems pairs are displayed in Figure 2. The second problem in each problem pair or example-problem pair were identical. These second problems were labeled “Problem 2” in the problem-solving control and “Your Turn” problems in the example conditions. These sample “Your Turn” problems are displayed in Figure 3.

The three worked examples conditions varied in whether they were correct or incorrect as well as whether students were prompted to study correct concepts and procedures or errors (displayed or potential). Students in the *Correct Explanation* condition studied a fictitious students’ correct method for solving a quadratic equation, using one of the four targeted techniques. These examples were paired with prompts that asked students to explain a correct concept or procedure displayed. The *Displayed Error* condition displayed the same problems as those in the *Correct Explanation* condition but instead included a common error that they were then prompted to explain. The *Potential Error* condition displayed the same correct examples as those displayed in the *Correct Explanation* condition, but students were instead prompted to explain a specific *potential* error that could occur in that particular problem. A visual of how explanation prompts would vary given the same target content are displayed in Figure 2. We provide another example in text for clarity.

For example, in an example that targets solving a quadratic equation through factoring, one problem solved within the example conditions was  $x^2 + 8x - 48 = 0$ . In the *Correct Explanation* condition, students are shown the correct factoring procedure and the solutions for this equation which are  $x = -12$  or  $x = 4$ . Then, students are asked “*When factoring, why did Himanshu use 12 and -4 as factors?*”. In the *Displayed Error* condition students are shown the fictitious student using the wrong factors and their prompt reads, “*When factoring, Himanshu did*

*not use the correct factors of -48. Why can't Himanshu use 24 and -2?*". In the *Potential Error* condition students are shown the correct solution and then asked about a potential error, "*When factoring, why can't Himanshu use 24 and -2 as factors instead?*". Thus, the explanation prompts across these conditions all focused students' attention on finding the appropriate factors for the same quadratic equations. What varied about the examples and their corresponding prompts here was whether students were shown the correct strategy and solution (i.e., Correct Explanation, Potential Error) or the factoring error (i.e., Displayed Error) and whether students were asked to reflect on the correct factoring (i.e., Correct Explanation) or incorrect factoring (i.e., Potential Error, Displayed Error). In the Problem-solving Control group, students are asked to solve the equation through factoring. Here we provide several sample examples, prompts, and target concepts and errors. For a full list of the target concepts and errors in all of the experimental assignments, readers are encouraged to refer to Barbieri and Booth (2020).

## ***Procedure***

### *Data Collection*

The current study was exempt from informed consent as the procedures involved were considered normal classroom practices by the IRB at XX University. Participants were randomly assigned within classroom to one of four experimental conditions: three worked examples conditions and one problem-solving control. The study followed a pretest–experimental manipulation–posttest design. After pretesting, regular classroom teachers administered the four assignments over the course of several weeks as the class completed their unit on solving quadratic equations. Students worked independently on assignments and teachers did not discuss differences between the assignments amongst the class. After completion of the fourth worksheet, students then completed a posttest which was identical to the pretest. Teachers used

the assignments as they normally would use other assignments, which was typically as practice after whole class instruction. Students in all conditions were allowed to ask their teachers questions while working with the materials. However, no systematic feedback was provided on any problems solved within the control group or on the example worksheets. Full data collection procedures are summarized in Barbieri and Booth (2020). The current analysis focuses on process data so that is the focus of the procedures described in this report.

### *Error Coding Procedure*

We devised an error coding scheme to help expound the prevalence of conceptual errors that participants made on four assignments completed between pretests and posttests. There were two worked example conditions that promoted error study and/or explanation (i.e., *Displayed Error*, *Potential Error*) and two conditions did not (i.e., *Correct Explanation*, *Problem-solving Control*). Each example condition included twelve example-problem pairs, so students completed a total of 12 practice problems. Within each problem, there were two general types of errors that could possibly be made on each problem: conceptual errors that represent targeted concepts within the examples and arithmetic errors. The error coding scheme was designed to code conceptual targeted errors. We term them conceptual in nature because they focus on the (mis)understanding of a particular mathematical principle or feature that could likely lead to the error displayed, rather than a simple arithmetic error. Arithmetic errors were not targeted in the current coding scheme or study as they were not the focus within the errorful learning conditions (i.e., *Displayed Error* and *Potential Error*). We designed this coding scheme to create a measure of conceptual error prevalence. We coded different errors for 12 different problems (three problems within each of the four worked example assignments) which were then used to create a composite measure of conceptual errors.

Every topic had different target errors that commonly arise while learning those topics (see Table 1 for examples). A total of forty-one errors were considered targeted conceptual errors. Errors were coded as “1” if the student displayed the target error or “0” if the student did not make the targeted error. Each problem was also coded to determine whether the question was attempted and solved correctly for descriptive purposes. Table 2 provides descriptive statistics by condition. To calculate the proportion of errors made by students in different conditions while solving quadratic equations on problems paired with each example, we calculated the mean of conceptual errors made across all assignments. The second problem within each problem-set in the control group was identical to the “Your Turn” problems paired with the examples. Thus, conceptual errors were also targeted on these particular problems within the control group for comparison purposes. This operationalization of misconception—as error types of a conceptual nature that are frequently made even after being directly targeted—echoes that which was used by Barbieri et al. (2019). If a specific condition yields a lower prevalence of conceptual errors than another condition, and error prevalence predicts posttest, this may reveal an underlying mechanism of the different learning benefits found in the prior analysis by condition (Barbieri & Booth, 2020). All errors were coded by two independent coders. To assess agreement amongst the two coders, we calculated mean differences between the two sets of codes and also calculated bivariate correlations between the two sets of codes. The mean difference between coders was 1.01% for targeted errors suggesting a high level of agreement. The correlation between the two coders was strong at  $.77$  ( $p < .001$ ).

#### *Explanation Coding Procedure*

Each worked example was paired with at least one self-explanation prompt to which students responded. If a specific condition evokes a different type of explanation than another condition at

a significantly higher rate, and that explanation type predicts posttest, this may reveal an underlying mechanism of the different learning benefits found in the prior analysis by condition (Barbieri & Booth, 2020). To examine this, we created a coding scheme devised to describe and quantify the relative proportion of student explanations that were conceptual versus procedural in nature with the goal of comparing whether the three example conditions varied in the type of thinking they evoked from students (i.e., differential focus on concepts vs. procedures). Explanations were also coded for correctness, attempts, and ambiguity for descriptive purposes. The first coder coded explanations for the entire sample. A second coder coded 30% of the sample. To assess agreement among the two coders, we calculated mean differences between conceptual and procedural explanations and also calculated bivariate correlations between the two sets of codes. The mean difference between coders was 0.31% for procedural explanations and 3.79% for conceptual explanations suggesting a high level of agreement. Correlations between the two coders were  $.72$  ( $p < .001$ ) for conceptual explanations and  $.63$  ( $p < .001$ ) for procedural explanations suggesting an acceptable level of consistency in explanation coding. Sample codes for one explanation are in Figure 4. Descriptive statistics by condition are presented in Table 2.

## Results

In the prior study (i.e., Barbieri & Booth, 2020), we found that the three worked examples conditions outperformed the *Problem-solving Control* on the algebraic feature knowledge measure. We also found that the *Displayed Error* condition in particular outperformed the *Problem-solving Control* on the equation-solving posttest. First, we review findings and corresponding analyses from the prior study (Barbieri & Booth, 2020) in more depth below. Then, we proceed to present the data analytic plan for the current study as well as

results, which examined the underlying mechanisms of benefits on algebra performance by example type.

### ***Summary of Prior Analyses and Results***

Missing data analysis was conducted and reported in Barbieri and Booth (2020) and revealed rates of missingness on posttests were high at 24.8%. This high attrition rate was due to several factors that normally arise in classroom research (i.e., teachers or students absent, students pulled out for services, school delays due to inclement weather). Missingness rates did not vary by condition. As posttests were likely not completed due to logistical reasons (e.g., substitute coverage) and not due to students' potential posttest scores had they completed the study, we assume data are missing at random (MAR). MAR is appropriately addressed through Full Information Maximum Likelihood. FIML is the gold standard for handling missing data (see Allison, 2003) FIML uses all information available in the dataset to estimate parameters without requiring imputation or deleting participants with missing values (Acock, 2012).

Analyses in the prior study were conducted through a series of path analyses with observed variables in MPlus Version 7.3 (Muthén & Muthén, 1998 - 2014). Three dummy-coded variables were created to include condition in models with *Problem-solving Control* as the referent group. Differential effects by condition were assessed in path models of manifest variables (Table 3). There were no differences at pretest between conditions so baseline balance was established (Table 4). All three worked examples conditions outperformed the *Problem-solving Control (Control)* group on posttest algebraic feature knowledge. However, only the *Displayed Error* condition outperformed the *Control* group on posttest equation-solving. Thus, there is a unique yet modest benefit of studying and explaining displayed errors within worked examples on equation-solving skills. Yet all forms of worked examples tested are significantly

more and equally effective at improving algebraic feature knowledge in comparison to the *Control* condition alone. We briefly summarize and present these findings (Table 4 and Figure 5), originally published in Barbieri and Booth (2020), to display the rationale for the current follow-up analyses.

### ***Data Analytic Plan for the Current Study***

We propose two mechanisms that explain the differential main effects on equation-solving and feature knowledge as seen in Barbieri and Booth (2020). For feature knowledge, we expect that the three worked examples conditions may have reduced misconception prevalence to a greater extent than the *Problem-solving Control* which could help to explain the marked improvements in feature knowledge. For equation-solving, we expect that the *Displayed Error* condition increased learners' focus on specific problem-solving procedures more so than the *Control* and that this emphasis helps to explain this condition's particular benefit on equation-solving skills at posttest. Thus, we tested two mediation models: one focused on assessing the indirect effects of condition on feature knowledge through misconception prevalence (Aim 1) and the other assessing the indirect effects of condition on equation-solving through increased emphasis on procedural knowledge (Aim 2).

We operationalized each of the mechanisms assessed through in-depth coding of process data, as described in the methods section. To briefly review, we operationalize misconception prevalence as the proportion of targeted conceptual errors students made on the problems they solved that were paired with each example that targeted that corresponding error, similar to Barbieri and colleagues (2019). More specifically, we coded whether the student made the particular error(s) being targeted or addressed within that example on the problem that the example was paired with. We then created a composite score for each student that represented

the proportion of targeted errors made of those possible to make across all examples. To operationalize increased attention or focus on algebraic procedures, we coded the explanations students provided on the self-explanation prompts paired with each example. More specifically, we coded whether each explanation was more conceptual or procedural in nature. We then created composite scores for each student that represented the proportion of explanations provided that were more conceptually-focused and the proportion of explanations provided that were more procedurally-focused. First, we assessed the indirect effects of the three worked examples conditions on algebraic feature knowledge posttest scores through the potential mediator *misconception prevalence*, with problem-solving control as the referent. That is, we ran a path model with manifest variables and an indirect effect to assess the mediation.

In the first model, three dummy-coded variables were used to represent each of the worked examples conditions (i.e., *Displayed Error*, *Correct Explanation*, *Potential Error*) making *Problem-solving Control* the referent. Exogenous (i.e., independent) variables (which here include the three worked example condition dummy codes) are automatically correlated in path analysis which ensures that the relationships between predictors are also accounted for through simultaneous estimation. This model is displayed in Figure 6.

Next, we assessed the indirect effects of the *Displayed Error* condition on equation-solving posttest scores through the potential mediator *procedural focus*. To operationalize increased attention or focus on algebraic procedures, we provided whether explanations students provided on self-explanation prompts paired with each example were more conceptual or procedural in nature. For this particular model we were unable to use the control group as a referent because students in the control condition were not asked to provide explanations. Thus, we use only one dummy code for the *Displayed Error* condition and the two other example-

based conditions (i.e., *Potential Error and Correct Explanation*) serve as the referent group. We acknowledge that this is a limitation in the current analysis and discuss the implications of this in the discussion section. This model is displayed in Figure 7.

An indirect effect is a measure of the effect of a predictor (e.g., condition) on an outcome (e.g., posttest scores) through a mediator (e.g., misconception prevalence, procedural focus). As indirect effects are estimated by calculating the product of the direct effects (typically less than one) that comprise them, they are typically smaller in magnitude than either direct effect alone. Because indirect effects are a product term of regression coefficients, they may violate the assumption of normality. Thus, we assessed the indirect effect with the classic Sobel test (1982) and used Preacher and Hayes (2004) bootstrapping procedure to adjust standard errors (advised by MacKinnon et al., 2004). Standard errors are estimated by taking repeated random samples from the existing data (in our case 10,000) and requires fewer assumptions about the normality of the distribution of the indirect effect. These bootstrapped samples assist in the handling of missing data.

***Aim 1: Indirect Effect of Worked Examples on Posttest Feature Knowledge through Decreased Misconception Prevalence***

First, we tested a model assessing whether a change in misconception prevalence mediated the relationship between condition and algebraic feature knowledge. According to Barbieri and Booth (2020), the three worked examples conditions outperformed the *Problem-solving Control* group at posttest on the algebraic feature knowledge measure. Conditions must vary on misconception prevalence for this measure to be a potential mechanism. Misconception prevalence was operationalized as the proportion of targeted conceptual errors made on the “Your-Turn” problems. Thus, we first established that conditions did vary on misconception

prevalence ( $F[3,188] = 4.40, p = .005$ ), with all three example conditions making significantly fewer targeted errors than the *Control*. Next, to determine whether this lowered misconception prevalence accounted for some of the benefits found at posttest, we fit a path model and tested the indirect effects of all three example conditions (three dummy codes with *Problem-solving Control* as reference) on posttest feature knowledge through misconception prevalence, controlling for pretest algebra feature knowledge and pretest equation-solving. Path coefficients are interpreted similarly to regression coefficients in multiple regression. Table 5 and Figure 6 display several paths crucial to measuring indirect effects. There are significant indirect effects of each example condition on posttest algebraic feature knowledge through misconceptions (i.e., targeted errors). The *Displayed Error*, *Potential Error*, and *Correct Explanation* conditions each made fewer targeted errors than the *Control* which then predicted higher algebraic feature knowledge scores at posttest ( $\beta = .10^*$ ;  $\beta = .15$ ;  $\beta = .13$ , respectively,  $ps < .05$ ). Therefore, the lower misconception prevalence seen in the worked example conditions compared to the control explained *some* of the increased algebra feature knowledge found at posttest.

***Aim 2: Indirect Effect of Displayed Error Condition on Posttest Equation-solving through Increased Procedural Focus***

According to Barbieri and Booth (2020), the *Displayed error* condition outperformed the *Problem-solving Control* group at posttest on the equation-solving measure. Conditions must vary on procedural focus for this measure to be a potential mechanism. Procedural focus was operationalized as the proportion of student explanations of the worked examples that were procedurally-focused in nature. Thus, we first established that conditions did vary on procedurally and conceptually-focused explanations ( $F[2,133] = 3.05, p = .051$ ;  $F[2,133] = 6.29, p = .002$ , respectively). The ANOVA comparing procedurally-based explanations by example-

condition was marginally significant (i.e.,  $p = .051$ ) and suggested that the *Displayed error* condition provided marginally more procedurally-focused explanations than the other two example conditions. The *Displayed Error* condition provided significantly less conceptually-focused explanations than the *Potential Error* condition but did not differ from the *Correct explanation* condition (see Table 2 for descriptives and comparisons). We then tested a model assessing whether this potentially increased procedural focus in students' explanations mediated the relationship between the *Displayed error* condition and equation-solving at posttest. That is, to determine whether this slightly higher proportion of procedural explanations accounted for some of the benefits found at posttest, we fit a path model and tested the indirect effects of the *Displayed Error* condition on posttest equation-solving through procedural focus, controlling for pretest equation-solving. Table 6 and Figure 7 display several paths crucial to measuring indirect effects. We found a marginally significant indirect effect of the *Displayed Error* condition on equation-solving posttest through procedural explanations with a  $p$ -value of just over the standard  $\alpha = .05$  ( $p = .054$ ). That is, the *Displayed error* condition tended to yield more procedural explanations than the *Potential error* and *Correct explanation* conditions which then predicted higher equation-solving scores at posttest ( $\beta = .06$ ). However, it is important to note that this indirect effect was small in magnitude. Implications are considered in the discussion.

## Discussion

Our current findings reveal that worked examples reduce misconceptions regardless of whether they are in the form of correct or incorrect examples. This reduction in misconceptions explained greater improvements in students' algebraic feature knowledge compared to a problem-solving control. It is interesting to consider that showing students either a correct way to solve a problem or a common error students generally make for that problem type reduces the

rate at which they make that type of error on similar problems. Thus, it seems that either strengthening correct knowledge or increasing negative knowledge (i.e., what *doesn't* work, what *not* to do) seem to weaken incorrect knowledge with approximately equivalent effects. However, it is important to note that there was some variability in the effects with the *Displayed Error* condition (i.e., incorrect examples paired with prompts for explaining the error) having the largest impact ( $\beta = -.275$ ) with very similar effects from the *Correct Explanation* condition (i.e., correct examples paired with prompts for the correct concept or procedure;  $\beta = -.244$ ). The smallest effect was yielded by the *Potential Error* condition (i.e., correct examples paired with prompts asking students to anticipate a common error for the problem type;  $\beta = -.187$ ). It may be slightly more challenging for students to study a correct example and simultaneously consider a potential error without actually seeing one, which may explain the dampened (though still significant) effect on reduction of misconceptions. Therefore, if choosing between implementing example types in the classroom, it seems that using either correct examples or incorrect examples paired with prompts that focus on the consistent form of explanation may be prudent.

One key finding to highlight is that having students study and explain errors within incorrect examples did *not* increase the likelihood of students making those errors. In fact, studying and explaining errors *or* studying correct examples and explaining potential errors led to fewer errors than the problem-solving control and just as few errors as the correct example condition. This is a critical point because practitioners may be hesitant to display errors within their classrooms for fear that students may adopt some of the incorrect strategies that they are asked to study. Our findings demonstrate that this fear is unfounded and that studying and explaining errors is one effective way to reduce problem-solving errors compared to simple problem-solving practice.

Though all three example types show equivalent benefits on reducing the prevalence of making common errors, there are indeed benefits specific to incorrect examples. Students who studied and explained incorrect examples showed significantly greater improvements in their equation-solving skills which was partially evoked through focusing students' attention on procedurally-based understandings as evidenced by their increased prevalence of explanations that focused on problem-solving steps. Of note is that conceptually-focused explanations did not predict equation-solving. The benefit of incorrect examples seems to be somewhat due to the increase in procedurally-focused explanations. Explaining why an error is wrong can help learners identify the particular features of the problem that make the solution incorrect and that seems to lead to strengthening of correct strategies and refinement of problem-solving skills.

Findings from Barbieri and Booth (2020) revealed that all three worked examples conditions lead to greater improvements in algebraic feature knowledge than the *Problem-solving control* but only the *Displayed Error* condition (i.e., studying and explaining incorrect examples) outperformed the *Control* in equation-solving. Yet the mechanisms explaining these benefits were unclear. Though prior work suggests reasons for varying impacts of example-types, the current study is the first to specifically assess mechanisms of learning from errors in particular.

One alternative mechanism not assessed in the current study that could at least partially explain the overall example effect is that although no systematic feedback was given to students in any condition, information presented within the example materials could have served as a form of indirect feedback on students' problem-solving on the "Your Turn" problems. Students in the control group solved problems and did not have access to additional information in the materials that may have provided indirect feedback. As our control group did not provide

feedback, we cannot directly compare whether a problem-solving control group with feedback would yield the same effects as worked examples without direct feedback. However, we would like to note that this possible alternative explanation does not truly oppose some of the mechanisms we propose in the current study. That is, we focus on potential mechanisms of a reduction of misconceptions (i.e., fewer targeted conceptual errors), and an increase in principled algebra knowledge (i.e., explanations focusing on principles underlying procedures). It could be that the information provided within the examples served as indirect feedback which then led to a reduction of misconceptions and an increased focus on principles underlying procedures. This possibility does not negate our findings but merely presents a more nuanced explanation of the possible mechanisms underlying the effects found. Future work comparing a control group with feedback to example conditions without feedback could clarify this potential addition to the proposed mechanisms. One past study examining the differential effects of worked example type (Durkin & Rittle-Johnson, 2012) found that students who studied a combination of correct and incorrect examples discussed more correct concepts than those who studied solely correct examples, and that discussing correct concepts predicted procedural knowledge. The authors suggested that incorrect examples may benefit learners by provoking greater frequency of discussing correct concepts. However, they tested this with an ANCOVA model, by using discussion of correct concepts as a covariate or predictor of posttest conceptual and procedural knowledge, rather than as a mediator of the effects of condition found. Our current analysis of the process data from Barbieri and Booth (2020) provides a more comprehensive picture of how learners benefit from varying example types by testing mechanisms through mediator models. It may seem that our findings conflict with that of Durkin and Rittle-Johnson (2012), since we did not find the prevalence of conceptual explanations as a significant predictor of equation-solving

skills. However, it is important to note that our coding procedures for explanations were different from Durkin and Rittle-Johnson (2012). Although we coded correct explanations as either conceptual or procedural in nature and came up with distinct scores for each, Durkin and Rittle-Johnson (2012) coded either comparison-based explanations or conceptual explanations (that did not include comparison). That is, they did not create a separate coding procedure for procedurally-based explanations without comparison. Additionally, some examples provided for the conceptual explanations may also be interpreted as procedural according to the nature of our coding scheme. Thus, it is possible that, had they separated out explanations into a similar dichotomy as that used in the current study, they may have found similar results.

There are several methodological strengths of the current study. First, random assignment of students to condition within classrooms ensures that we can make causal claims for condition effects. In addition to random assignment, we were also able to establish baseline balance which ensures that students across conditions were equivalent on key study measures. The inclusion of a problem-solving control as well as varying types of example conditions enabled us to test not only an overall worked examples effect compared to the control but also in comparing example-types to each other. Further, and most importantly, we captured not only pre-posttest change but also leveraged rich process data (i.e., student explanations and problem-solving errors) to gain a deeper understanding of learning processes that occur while working with the study materials. Explanation prompts in the example conditions were designed to target the same concepts or procedures regardless of whether they were presented within a correct or incorrect example, ensuring that what varied amongst the example conditions was simply correctness of example rather than difference in target content knowledge. Further, students completed the same practice problems in all conditions, making direct comparison of problem-solving errors possible. Our

analytic approach also has several strengths. First, we used a gold-standard approach for handling missing data (FIML) ensuring that no student was dropped from analyses regardless of whether they missed a day of data collection. We also used simultaneous estimation of effects within our path models, thereby reducing the risk of Type I error which can be a particular issue in mediation analyses run in a stepwise fashion. Finally, we paired the Sobel test with bootstrapped standard errors to address potential issues with non-normality of indirect effects.

Alongside these strengths are several limitations of the current study. First, our problem-solving control group was not prompted to provide explanations. Thus, we were unable to compare the role of explanation-type within each of the example-conditions to the control. Rather, we used the *Correct Explanation* and *Potential Error* conditions as the referent group when examining the role of example-type in improvements yielded on equation-solving through the *Displayed Error* condition. Though informative, this analysis is not perfectly analogous to the direct effects model of equation-solving tested in Barbieri and Booth (2020). That is, in the overarching study, the *Displayed Error* condition outperformed the *Problem-solving Control*, but *not* the two other example conditions, on equation-solving at posttest. The *Correct Explanation* and *Potential Error* conditions did not differ at posttest on equation-solving from either the *Displayed Error* or control conditions. This particular limitation could explain why the indirect effect of the *Displayed Error* condition on posttest equation-solving through procedurally-focused explanations was marginally significant and just over the standard threshold of  $\alpha = .05$  (i.e.,  $p = .051$ ). If we were able to use the *Problem-solving Control* as a referent rather than the other example conditions, we may have seen a larger indirect effect with a smaller  $p$ -value. However, as the self-explanation prompts asked students to explain something displayed in the worked examples themselves, there was no seamless or logical way to include the same

explanation prompts in the control group materials. Yet we acknowledge that some explanation could have been prompted to help isolate the effects of examples themselves. However, we do not believe that the self-explanation prompts alone are what is driving the effects of the example-based conditions due to findings from a recent meta-analysis (Barbieri et al., 2023). This meta-analysis on the effects of worked examples on mathematics performance demonstrated a negative effect of pairing examples with self-explanation prompts. That is, although worked examples had a moderate positive effect ( $g = .48$ ) on mathematics performance measures, those studies with example conditions that included self-explanation prompts had smaller effects ( $\beta = -0.24^*$ ) than those that did not pair examples with explanation prompts. Thus, it is possible that the differences between example conditions and the problem-solving control would have been even more pronounced had we not included explanation prompts in the current experimental manipulation. This is an open question that can be addressed in future work. Further work is needed to examine the specific benefits of studying and explaining errors on equation-solving compared to a problem-solving control in particular. However, findings from the current study highlight an important mechanism to examine in such work.

Another limitation of our current methods includes our measure of misconceptions. That is, we operationalized misconceptions by quantifying the proportion of errors made that were directly targeted in examples after studying such examples. This is aligned with the method used by Barbieri et al. (2019) and what was possible with the data collected in the overarching study presented in Barbieri and Booth (2020). However, future work should include other measures of misconceptions to gain a richer measure of students' misunderstandings or misconceptions, such as interview or think-aloud data. We initially expected incorrect examples to reduce misconceptions to a greater extent than correct examples, as is suggested by prior work (Barbieri

& Booth, 2020; Metcalfe, 2017). Yet students who studied and explained incorrect examples made as few targeted errors as those who studied and explained correct examples. Richer measures may also illuminate the reason for this unexpected finding.

One important consideration to keep in mind is that the current analyses revealed partial mediations. That is, the proposed mechanisms did not fully explain the benefits found in Barbieri and Booth (2020). Future work is needed to examine what other mechanisms may explain the benefits of worked examples on algebraic feature knowledge as well as the benefits of incorrect worked examples of equation-solving skills. Many mechanisms of worked examples more generally have been proposed (see Sweller et al., 2019 for a review) such as reductions in cognitive load yet many of these proposed mechanisms have not been tested empirically with middle school students on algebraic feature knowledge in particular. Thus, there is much room for further exploration of these specific relationships. For incorrect examples in particular, these other mechanisms may include changes in varying forms of students' mathematical cognitions such as improvements in negative knowledge (e.g., Barbieri et al., 2021; Gartmeier et al., 2008) and improvements in metacomprehension (e.g., calibration) to name a few. There also may be particular attitudinal benefits from studying errors in particular in the context of incorrect examples. For example, using errors as learning tools in the classroom may alter how students perceive their own mistakes made in the mathematics classroom. This improvement in the general perception of the usefulness of errors (e.g., perceived functionality of errors; Steuer et al., 2013) may support students' uptake and use of errorful learning experiences that could promote larger effects of such materials. Barbieri and Booth (2016) suggested this potential mechanism but did not empirically test it. Thus, it is likely that there are many potential

mechanisms working together to bring about the changes found in Barbieri and Booth (2020) and in the current analyses as well.

Despite the limitations described above and the room for further exploration of mechanisms, the current study confirms that all example-types are beneficial for students' algebra learning, with varying effects and mechanisms by outcome. All three example-types were effective in improving algebra students' feature knowledge and studying and explaining incorrect examples was particularly effective at improving equation-solving skills. We now know that worked examples in general can effectively reduce common student misconceptions about algebra and that reduction supports their developing feature knowledge. Further, incorrect examples are particularly helpful in improving equation-solving skills because they focus students' attention on the procedures within the examples and strengthen students' knowledge of correct problem-solving strategies. These findings may inform the target content teachers emphasize in instruction as well as their interpretation of students' understanding. Knowledge that worked examples promote general conceptual understanding and incorrect worked examples in particular draw attention to problem-solving procedures may help teachers select which example type to use, given students' current level of understanding of algebra. Thus, incorporating worked examples, and specifically incorrect worked examples, into instruction has specific educational benefits for learners.

## References

- Acock, A. C. (2012). What to do about missing values. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 3: Data analysis and research publication*. (pp. 27–50). American Psychological Association. <https://doi.org/10.1037/13621-002>
- Adams, D. M., McLaren, B. M., Durkin, K., Mayer, R. E., Rittle-Johnson, B., Isotani, S., & van Velsen, M. (2014). Using erroneous examples to improve mathematics learning with a web-based tutoring system. *Computers in Human Behavior*, *36*, 401–411. <https://doi.org/10.1016/j.chb.2014.03.053>
- Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, *112*(4), 545–557. <https://doi.org/10.1037/0021-843X.112.4.545>
- Barbieri, C., & Booth, J. L. (2016). Support for struggling students in algebra: Contributions of incorrect worked examples. *Learning and Individual Differences*, *48*, 36–44. <https://doi.org/10.1016/j.lindif.2016.04.001>
- Barbieri, C. A., & Booth, J. L. (2020). Mistakes on display: Incorrect examples refine equation solving and algebraic feature knowledge. *Applied Cognitive Psychology*, *34*(4), 862–878. <https://doi.org/10.1002/acp.3663>
- Barbieri, C. A., Booth, J. L., Begolli, K. N., & McCann, N. (2021). The effect of worked examples on student learning and error anticipation in algebra. *Instructional Science*, *49*(4), 419-439.
- Barbieri, C. A., Miller-Cotto, D., & Booth, J. L. (2019). Lessening the Load of Misconceptions: Design-Based Principles for Algebra Learning. *Journal of the Learning Sciences*, *28*(3), 381–

417. <https://doi.org/10.1080/10508406.2019.1573428> Barbieri, C. A., Miller-Cotto, D., Clerjuste, S. N., & Chawla, K. (2023). A Meta-analysis of the Worked Examples Effect on Mathematics Performance. *Educational Psychology Review*, 35(1), 11.

Booth, J. L., Cooper, L. A., Donovan, M. S., Huyghe, A., Koedinger, K. R., & Paré-Blagoev, E. J. (2015c). Design-Based Research Within the Constraints of Practice: AlgebraByExample. *Journal of Education for Students Placed at Risk (JESPAR)*, 20(1–2), 79–100.

<https://doi.org/10.1080/10824669.2014.986674>

Booth, J. L., Oyer, M. H., Paré-Blagoev, E. J., Elliot, A. J., Barbieri, C., Augustine, A., & Koedinger, K. R. (2015b). Learning Algebra by Example in Real-World Classrooms. *Journal of Research on Educational Effectiveness*, 8(4), 530–551. <https://doi.org/10.1080/19345747.2015.1055636>

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last.

*Cognition*, 106(2), 1047–1058. <https://dx.doi.org/10.1016/j.cognition.2007.04.010>

Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review*, 34(4), 344–377. <https://doi.org/10.1016/j.dr.2014.10.001>

D’Mello, S., & Graesser, A. (2014). Confusion and its dynamics during device comprehension with breakdown scenarios. *Acta Psychologica*, 151, 106–116.

<https://doi.org/10.1016/j.actpsy.2014.06.005>

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction*, 22(3), 206–214.

<https://doi.org/10.1016/j.learninstruc.2011.11.001>

Gartmeier, M., Bauer, J., Gruber, H., & Heid, H. (2008). Negative Knowledge: Understanding Professional Learning and Expertise. *Vocations and Learning*, 1(2), 87–103.

<https://doi.org/10.1007/s12186-008-9006-1>

Gorsky, P., & Finegold, M. (1994). The role of anomaly and of cognitive dissonance in restructuring students' concepts of force. *Instructional Science*, 22(2), 75–90.

<https://doi.org/10.1007/BF00892158>

Heemsoth, T., & Heinze, A. (2014). The impact of incorrect examples on learning fractions: A field experiment with 6th grade students. *Instructional Science*, 42(4), 639–657.

<https://doi.org/10.1007/s11251-013-9302-5>

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence Limits for the Indirect Effect: Distribution of the Product and Resampling Methods. *Multivariate Behavioral Research*, 39(1), 99–128. [https://doi.org/10.1207/s15327906mbr3901\\_4](https://doi.org/10.1207/s15327906mbr3901_4)

McNeil, N. M. (2014). A Change-Resistance Account of Children's Difficulties Understanding Mathematical Equivalence. *Child Development Perspectives*, 8(1), 42–47.

<https://doi.org/10.1111/cdep.12062>

McNeil, N. M., & Alibali, M. W. (2000). Learning mathematics from procedural instruction: Externally imposed goals influence what is learned. *Journal of Educational Psychology*, 92(4), 734–744. <https://doi.org/10.1037/0022-0663.92.4.734>

Metcalfe, J. (2017). Learning from Errors. *Annual Review of Psychology*, 68(1), 465–489.

<https://doi.org/10.1146/annurev-psych-010416-044022>

Muthén, L. K., & Muthén, B. O. (1998-2014). Mplus user's guide (7th ed.). Los Angeles, CA: Muthén & Muthén.

Overoye, A. L., & Storm, B. C. (2015). Harnessing the power of uncertainty to enhance learning. *Translational Issues in Psychological Science*, 1(2), 140–148.

<https://doi.org/10.1037/tps0000022>

- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36(4), 717–731. <https://doi.org/10.3758/BF03206553>
- Resnick, L. B. (1983). Mathematics and Science Learning: A New Conception. *Science*, 220(4596), 477–478. <https://doi.org/10.1126/science.220.4596.477>
- Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a One-Way Street: Bidirectional Relations Between Procedural and Conceptual Knowledge of Mathematics. *Educational Psychology Review*, 27(4), 587–597. <https://doi.org/10.1007/s10648-015-9302-x>
- Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology*, 101(4), 836.
- Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York, NY: Oxford University Press.
- Sobel, M. E. (1982). Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. *Sociological Methodology*, 13, 290. <https://doi.org/10.2307/270723>
- Steuer, G., Rosentritt-Brunn, G., & Dresel, M. (2013). Dealing with errors in mathematics classrooms: Structure and relevance of perceived error climate. *Contemporary Educational Psychology*, 38(3), 196-210.
- Sweller, J. (1998). *From cognitive architecture to instructional design*. Paper presented at the Inaugural Seminar of Professor J.G. van Merriënboer, The Netherlands.
- Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261-292.

van Loon, M. H., Dunlosky, J., van Gog, T., van Merriënboer, J. J. G., & de Bruin, A. B. H. (2015).

Refutations in science texts lead to hypercorrection of misconceptions held with high confidence.

*Contemporary Educational Psychology*, 42, 39–48.

<https://doi.org/10.1016/j.cedpsych.2015.04.003>

**Tables and Figures****Table 1.*****Sample targeted errors***

Topic	Example of Targeted Error
Quadratic Formula	Failure to solve a quadratic equation for all solutions Not knowing how to put equation in standard quadratic form before trying to solve it
Factoring	Failure to set the factor to zero which is not aligned to the zero product property Not using the proper factors that add up and multiply to equal the coefficient of $b$ and $c$ in the standard quadratic equation $ax^2 + bx + c = 0$
Square Root	Trying to solve for the square root of a negative Failure to solve the negative square root of a number and instead only provide one solution based on the positive square root
Graphing	Thinking the $y$ -value of the vertex is the axis of symmetry (instead of the $x$ -value) Incorrectly graphing a quadratic function with a negative axis of symmetry as positive, with a parabola that is convex

**Table 2**

*Process data descriptives by condition*

<i>Measures</i>	Problem-solving		Correct		Displayed Error		Potential Error		Difference		Multiple
	Control		Explanation		Condition		Condition		Test by		Comparison
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>F</i>	<i>p</i>	<i>SNK</i>
Correct Your-Turn Problems (%)	32.60	25.95	59.87	26.40	52.29	23.68	59.59	25.19	12.72	<.001	PS < CE/DE/PE
Attempted Your-Turn Problems (%)	86.99	27.31	92.73	16.26	94.15	13.22	92.73	20.09	1.24	.296	<i>ns</i>
Targeted errors (Misconceptions) (%)	12.29	8.82	7.86	6.16	8.93	7.03	7.15	8.19	4.40	.005	PS > CE/DE/PE
Correct Explanations (%)	-	-	45.15	21.58	48.08	23.49	41.33	25.04	.949	.390	<i>ns</i>
Attempted Explanations (%)	-	-	82.09	30.69	84.04	27.35	80.14	33.49	.191	.827	<i>ns</i>
Procedural Explanations (%)	-	-	32.84	17.10	41.78	18.75	33.09	22.82	3.05	.051	<i>ns</i>
Conceptual Explanations (%)	-	-	26.01	16.49	23.36	15.25	14.65	15.98	6.29	.002	PE < DE/CE
Vague Explanations (%)	-	-	40.22	24.31	29.97	22.73	43.33	24.26	3.95	.022	DE < CE/PE
Answer Explanations (%)	-	-	1.35	3.18	5.94	9.32	5.39	9.72	4.45	.013	CE < DE/PE

*Note.* SNK = Student-Newman-Keuls posthoc comparisons; Because students in the Problem-solving control did not study worked examples, they were not prompted to provide explanations.

**Table 3**

*Maximum likelihood estimates of direct effects in main effects model of posttest scores (originally presented in Barbieri & Booth, 2020)*

Parameter	Post-Solving Quadratic Equations			Post-Algebraic Feature Knowledge		
	Unstandardized	SE	Standardized	Unstandardized	SE	Standardized
Pre-solving quadratic equations	.744**	.127	.486**	.477**	.101	.410**
Pre-algebraic feature knowledge	--	--	--	.045	.081	.040
Displayed error condition	12.052*	5.942	.200*	9.974*	4.682	.217*
Potential error condition	6.646	5.890	.112	9.813*	4.624	.216*
Correct explanation condition	2.895	5.781	.050	9.447*	4.537	.213*
Disturbance variances	473.797**	64.475	.730**	291.805**	39.772	.776**

*Note.* Estimates for Post-solving quadratic equations ↗ Post-algebraic feature knowledge: Unstandardized = 213.812\*\*, Standardized = .575\*\*.

Model fit:  $\chi^2_M [1] = .77, p = .378$ ; Comparative fit index [CFI] = 1.00; root mean square error of approximation [RMSEA] = .000 [.000 - .243];

Standardized root mean square residual [SRMR] = .014. \*  $p < .05$ , \*\*  $p < .01$

**Table 4**

*Establishing baseline balance: Descriptive statistics by condition (originally presented in Barbieri & Booth, 2020)*

	Problem-solving		Correct Explanation		Displayed Error		Potential Error		Difference Test by	
	Control		Condition		Condition		Condition		Condition	
	<i>n</i> = 50		<i>n</i> = 51		<i>n</i> = 52		<i>n</i> = 53			
<i>Continuous pretest measures</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>F</i>	<i>p</i>
Algebraic feature knowledge (%)	44.19	18.33	43.73	17.49	48.66	17.20	47.24	16.15	.971	.407
Solving quadratic equations (%)	15.65	14.18	17.78	18.51	15.64	14.15	18.78	18.23	.472	.702
<i>Categorical demographic variables</i>	%		%		%		%		$\chi^2$	<i>p</i>
Female	44.00		56.90		53.80		47.20		5.94	.115
URM	30.00		25.50		32.70		32.10		.780	.854
Low SES	16.00		7.80		15.40		17.00		2.23	.525

Note. URM = Underrepresented minority student; Low SES = received free or reduced lunch

**Table 5**

***Direct and indirect effects on Algebra Feature Knowledge Posttest***

	Feature Knowledge Model		
	Estimate	SE	p
<i>Direct Effects</i>	Algebra Feature Knowledge Posttest		
Algebra Equation-solving pretest	.29	.07	<.001
Algebra Feature Knowledge pretest	-.01	.09	.914
Correct Examples condition	.10	.11	.372
Potential Errors condition	.01	.11	.946
Displayed Error condition	.06	.11	.552
Targeted Errors (Misconceptions)	-.54	.09	<.001
<i>Direct Effects</i>	Targeted Errors (Misconceptions)		
Correct Examples condition	-.24	.08	.003
Potential Errors condition	-.19	.09	.006
Displayed Error condition	-.28	.10	.033
<i>Indirect Effects</i>	Algebra Feature Knowledge Posttest		
Correct Examples <i>through</i> Targeted Errors	.13	.05	.004
Potential Errors <i>through</i> Targeted Errors	.15	.06	.009
Displayed Errors <i>through</i> Targeted Errors	.10	.05	.039

Note. Problem-solving Control is referent group; Estimator: FIML

**Table 6**

*Direct and indirect effects on Algebra Equation Solving Posttest*

	Algebra Equation Solving Model		
	Estimate	SE	p
<i>Direct Effects</i>	Algebra Equation-solving Posttest		
Algebra Equation-solving pretest	.50	.07	<.001
Displayed Error condition	.05	.11	.615
Procedural Explanations	.31	.10	.002
Conceptual Explanations	.11	.10	.303
<i>Direct Effects</i>	Procedural Explanations		
Displayed Error condition	.20	.08	.009
	Conceptual Explanations		
Displayed Error condition	.08	.08	.330
<i>Indirect Effects</i>	Algebra Equation-solving Posttest		
Displayed Errors <i>through</i> Procedural Explanations	.06	.03	.054
Displayed Errors <i>through</i> Conceptual Explanations	.01	.01	.561

Note. Potential Error and Correct Example conditions as referent; Estimator: FIML

Figure 1.

Sample questions from the algebraic feature knowledge assessment and the equation-solving measure.

Algebraic Feature Knowledge Items	Equation-solving Items
<p>3. State whether each of the following is true for the quadratic function <math>y = -x^2 - 2x + 3</math>.</p> <p>a. The axis of symmetry is <math>x = 1</math>.                      Yes      No</p> <p>b. The axis of symmetry is <math>x = -1</math>.                      Yes      No</p> <p>c. The vertex is a minimum.                      Yes      No</p>	<p>1. Solve the quadratic equation by factoring. Show all of your work.</p> $0 = x^2 - 6x - 27$
<p>12. For each example (I – V), circle the correct next step (a or b) for solving using the quadratic formula</p> <p>D) <math>2x^2 - 3x + 2 = 0</math></p> $\text{Step 1: } x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ <p>Is Step 2 a or b?</p> <p>a. <math>x = \frac{-3 \pm \sqrt{(3)^2 - 4(2)(2)}}{2(2)}</math>                      b. <math>x = \frac{3 \pm \sqrt{(-3)^2 - 4(2)(2)}}{2(2)}</math></p>	<p>11. A ball falls off of a ledge. The equation <math>0 = -2t^2 + 10t + 0</math> gives the time <math>t</math> in seconds when the ball hits the ground (height of 0 feet). Use the quadratic equation to find out how many seconds it took for the ball to hit the ground.</p>

Figure 2. Sample questions from the (a) Correct explanation condition, (b) Displayed error condition, (c) Potential error condition, and (d) Problem-solving control. Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).

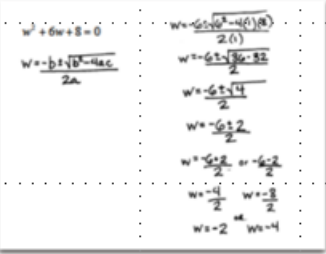
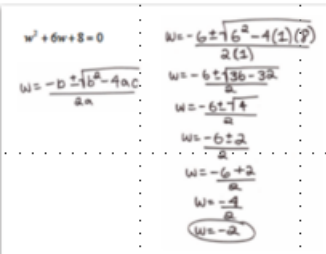
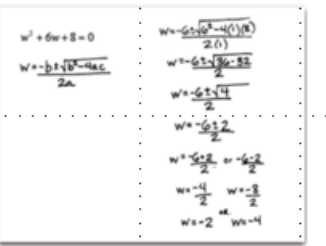
<b>Correct Explanation Condition</b>	
<p><input checked="" type="checkbox"/> Denzel solved this equation <b>correctly</b>. Here is his work.</p> 	<p><input type="checkbox"/> How did Denzel know to find two solutions?</p> <p><b>Your Turn:</b> <math>w^2 + 2w - 8 = 0</math></p>
<b>Displayed Error Condition</b>	
<p><input checked="" type="checkbox"/> Denzel <b>didn't</b> solve this equation correctly. Here is his work.</p> 	<p><input type="checkbox"/> There should be two solutions to this problem: What did Denzel forget to do to find the second solution?</p> <p><b>Your Turn:</b> <math>w^2 + 2w - 8 = 0</math></p>
<b>Potential Error Condition</b>	
<p><input checked="" type="checkbox"/> Denzel solved this equation <b>correctly</b>. Here is his work.</p> 	<p><input type="checkbox"/> If Denzel had just written <math>w = -2</math> as the solution, why would this be incorrect?</p> <p><b>Your Turn:</b> <math>w^2 + 2w - 8 = 0</math></p>
<b>Problem-solving Control</b>	
<p>1a. <math>w^2 + 6w + 8 = 0</math></p>	<p>1b. <math>w^2 + 2w - 8 = 0</math></p>

Figure 3.

## Sample “Your Turn” Problems by Targeted Approach


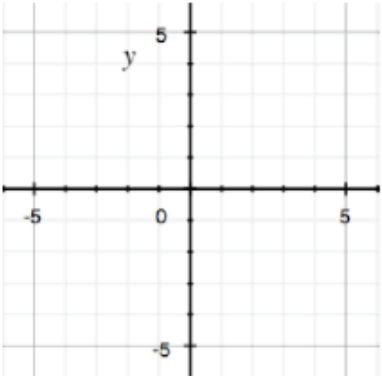



<p> <b>Your Turn:</b></p> $y = x^2 - 2x - 1$ <table border="1" data-bbox="331 618 703 821"> <tbody> <tr> <td>Axis of Symmetry</td> <td></td> </tr> <tr> <td>Vertex</td> <td></td> </tr> <tr> <td>Min or Max?</td> <td></td> </tr> </tbody> </table>  <p><i>Solving by graphing</i></p>	Axis of Symmetry		Vertex		Min or Max?		<p> <b>Your Turn:</b></p> $1 + 3x^2 = -5x$ <p><i>Solving with the quadratic formula</i></p> <hr/> <p> <b>Your Turn:</b></p> $x^2 + 3x - 28 = 0$ <p><i>Solving quadratics by factoring</i></p> <hr/> <p> <b>Your Turn:</b></p> $(n - 3)^2 = -4$ <p><i>Solving quadratics using the square root</i></p>
Axis of Symmetry							
Vertex							
Min or Max?							

Figure 4.

## Worked Example with Sample Explanation Codes



Abdalla **didn't** solve the equation correctly. Here is his work:

$$\begin{array}{l}
 4w^2 - 4w = -1 \\
 4w^2 - 4w - 1 \\
 w = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\
 w = \frac{4 \pm \sqrt{(-4)^2 - 4(4)(-1)}}{2(4)} \\
 w = \frac{4 \pm \sqrt{16 + 16}}{8} \\
 w = \frac{4 \pm \sqrt{32}}{8} \\
 w = \frac{4 + \sqrt{32}}{8} \quad \text{or} \quad w = \frac{4 - \sqrt{32}}{8} \\
 w = 1.21 \quad w = -.21
 \end{array}$$

What did Abdalla forget to do before applying the quadratic formula?

**Sample correct response:** Abdalla forgot to add 1 to both sides before applying the quadratic formula to set the function to zero. This puts it in standard form of a quadratic equation and makes the coefficient of c negative which is incorrect.

#### Explanation Type

- **Conceptual Response:** Demonstrates understanding that the function needs to be set to zero so that it is in standard form of a quadratic function
- **Procedural Response:** Talks about a specific procedure for solving equation without noting that setting the function to zero makes the coefficient of c positive
- **Answer Response:** Refers to an answer without explanation
- **Don't Know Response:** Reports lack of understanding
- **Vague Response:** Unclear responses that does not clearly indicate students' reasoning

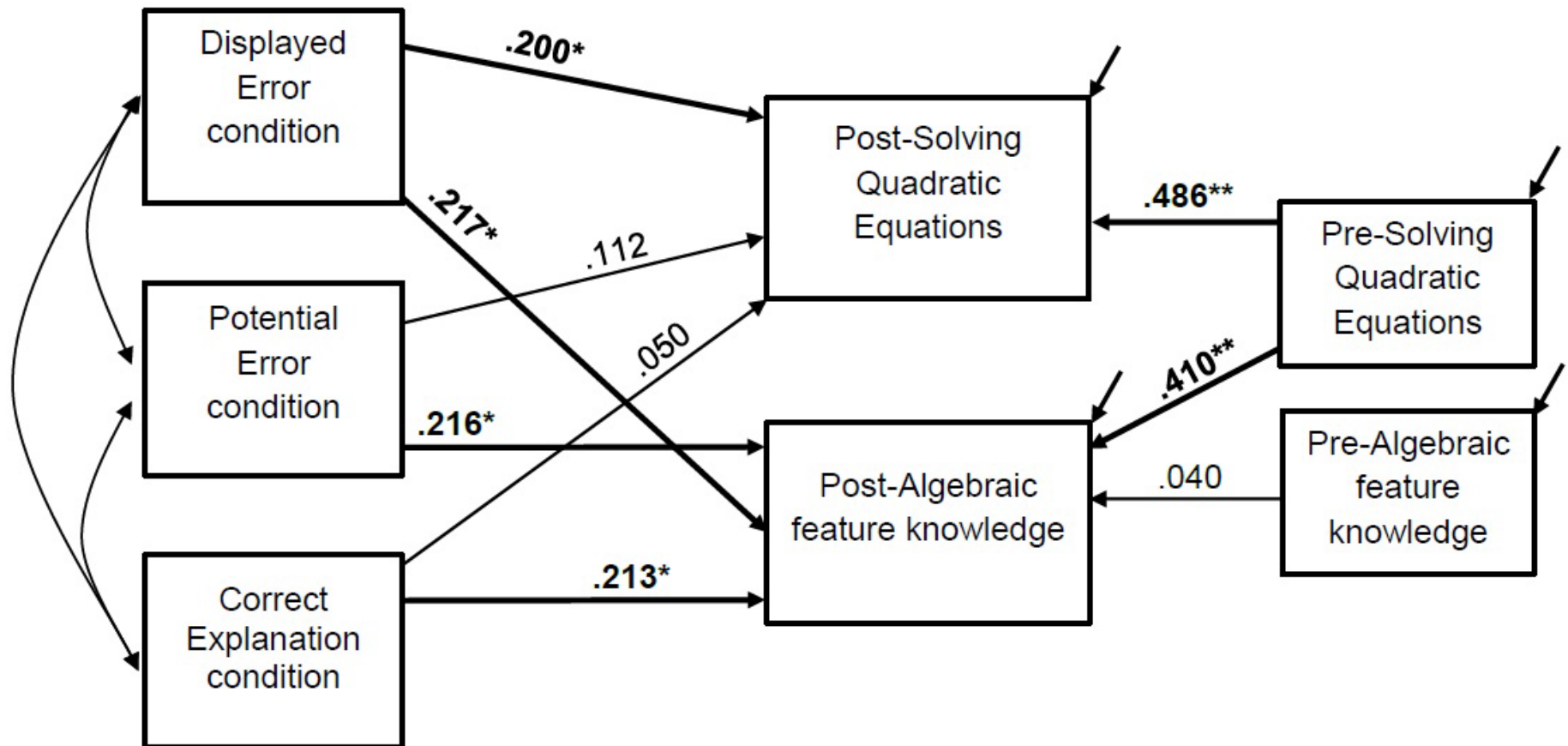


Figure 5. Main effects model predicting posttest scores of solving quadratic equations and algebraic feature knowledge (originally published in Barbieri & Booth, 2020).

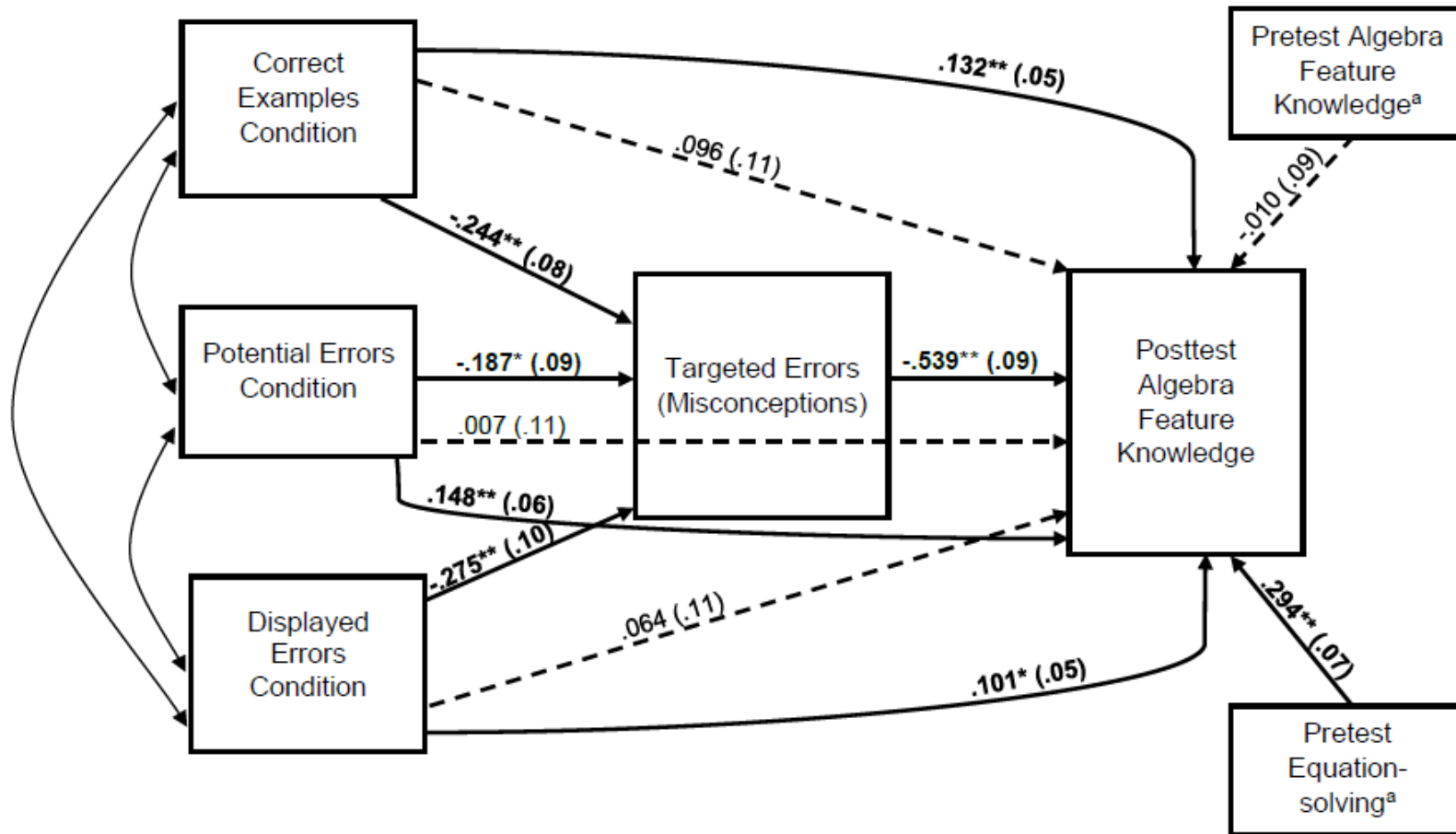


Figure 6. Path model displaying direct and indirect effects of conditions on posttest feature knowledge through misconceptions

*Note.* \*  $p < .05$ ; \*\*  $p < .001$ ; Curved line represents indirect effect of conditions on posttest scores through mediator (misconceptions). Dashed line represents nonsignificant path. <sup>a</sup> Pretest scores are also exogenous variables correlated with condition dummy codes. Curved double-arrows not shown for simplicity.

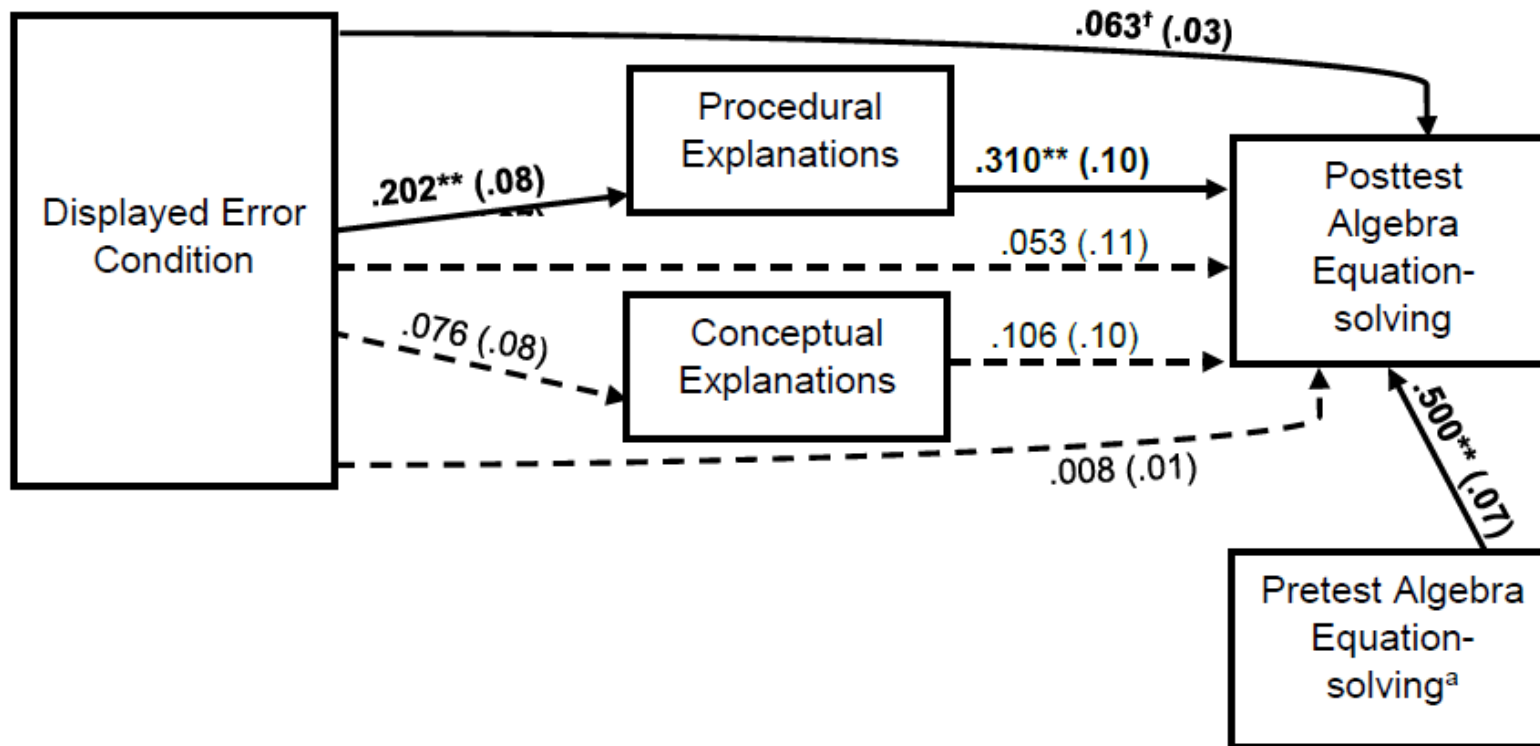


Figure 7. Path model displaying direct and indirect effects of Displayed Error condition on posttest equation-solving through explanation type

Note. \*  $p < .05$ ; \*\*  $p < .001$ ; †  $p = .051$ . Curved line represents indirect effect of condition on posttest scores through mediators (procedural and conceptual explanations). Dashed lines represent nonsignificant paths. <sup>a</sup>. Pretest score is also exogenous variable correlated with condition dummy code. Curved double-arrows not shown for simplicity.