USING ERROR ANTICIPATION EXERCISES AS AN INSTRUCTIONAL INTERVENTION IN THE ALGEBRA CLASSROOM

A Dissertation
Submitted to
the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
DOCTOR OF PHILOSOPHY

---

by

Nicholas F. McCann
May 2019

Examining Committee Members:

Dr. Kristie J. Newton, Advisory Chair, Department of Teaching and Learning
Dr. Julie L. Booth, Department of Psychological Studies in Education
Dr. Meixia Ding, Department of Teaching and Learning
Dr. Kelly M. McGinn, External Member, Department of Psychological Studies in Education

ABSTRACT

Researchers and instructors have only recently embraced the role of errors as vehicles for learning in the algebra classroom. Studying a mixture of correct and incorrect worked examples has been shown to be beneficial relative to correct worked examples alone. This study examines the effectiveness of having students generate, or anticipate, errors another student might make. Five Algebra 1 sections at a suburban mid-Atlantic public high school participated amid an early equation-solving unit. During teacher-led instruction, all five sections examined 2-3 correct worked examples. The final example varied across conditions. One section received an additional correct worked example. Two sections examined an incorrect worked example. The remaining two sections engaged in an error anticipation exercise where the teacher wrote an equation on the board and asked the students to predict errors another student might make in solving. The study measured conceptual and procedural knowledge, encoding ability, and student-generated errors. Although no meaningful significant differences were found, students in the error anticipation condition saw no difference in performance in conceptual and procedural items versus those who examined incorrect worked examples. Analysis that combined the error anticipation and incorrect worked examples conditions showed that those students trended toward outperforming those who examined correct examples only on procedural items. These results support further examination of error anticipation as a worthwhile instructional activity.

ACKNOWLEDGEMENTS

This has been quite a journey, and there are a number of people to whom I owe thanks. First and foremost I would like to thank my advisor, Dr. Kristie Jones Newton, for her unyielding support. I would never have reached this point without her assistance and encouragement, and I am forever grateful.

I would like to thank my committee members, Dr. Julie Booth and Dr. Meixia Ding, for their time and attention throughout the process. I appreciate the ways they challenged me to become a better researcher, writer, and thinker. Thanks also to Dr. Kelly McGinn for serving as the external reader and offering very thoughtful feedback. I consider myself lucky to have worked with such a brilliant committee.

Thank you to my family, friends, and coworkers for your encouragement over the years. Most notably to my wife, Jennifer, for her support and understanding. I have asked much of her throughout this journey and she has never hesitated to do what is necessary to help me finish.

Finally a special thanks to my daughter, Courtney, for inspiring me to push on and complete the process. No title will ever mean more to me than being called your dad.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION TO THE STUDY

Statement of Problem

In the near future, the United States will face increased competition for international talent in science, technology, engineering, and mathematics (STEM) to fill its needs and stagnation in the number of American-born STEM graduates (National Science Board, 2003; 2012). These threats are often phrased in terms of not only their potentially-negative economic consequences, but also in their potential to affect our national security and quality of life overall (National Mathematics Advisory Panel, 2008). Both cases are exacerbated by the reality that young Americans are continuing to avoid postsecondary studies within the STEM fields. The percentage of US students who enter science and engineering programs is about half that of Europe and one-sixth that of China (National Academy of Sciences, 2007). As a result, recent increases in science and engineering graduates from the European Union (80%) and Asia (134%) dwarfs that of the US (26%) (National Science Foundation, 2007).

An aversion to highly technical postsecondary programs is not surprising given US students' relative shortcomings in mathematics using international comparisons. American 15-year-old students scored below average relative to other developed countries on the recent results from the Programme for International Student Assessment (OECD, 2010; PISA, 2009, 2012). On the 2011 administration of the Trends in International Mathematics and Science Study (TIMSS), United States students showed mixed results. Fourth-grade students displayed an increase in mathematics performance and scores for eighth graders stagnated while foreign students demonstrated improvement in both cohorts (Provasnick et. al, 2012).

As a result of these continued deficiencies, student performance in Algebra has been identified as a critical area of focus for researchers and policymakers (National Mathematics Advisory Panel, 2008). Algebra is considered the "gateway" to more advanced mathematics and science courses (U.S. Department of Education, 1997; Moses & Cobb, 2001). Students who pass Algebra by the end of 9th grade take more mathematics classes in high school and are more likely to attend college (Evan, Gray, & Olchefske, 2006).

Overview

Given the urgency to improve student performance in mathematics generally and algebra specifically, current research has been exploring a range of possible classroom tools. One major area of research involves the use of self-explanation of worked examples, where students in the classroom are presented with a problem and possible solution method and are asked to study and respond to prompts about it. The use of self-explanation of worked examples has been shown to increase student conceptual and procedural knowledge, particularly within the Algebra domain (Sweller & Cooper, 1985; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Booth & Koedinger, 2008).

These exercises, although effective, are further enhanced with the inclusion of incorrect worked examples **(**Booth and Cooper et al., 2015; Booth and Oyer et al., 2015). Including incorrect examples has found increasing favor from researchers and practitioners as a viable way to address common misconceptions for students who struggle. Although the role of errors in furthering student knowledge also has a long history of research (Radatz, 1980), their day-to-day use in the classroom did not find favor until the 1990s ushered in reform-based mathematics curricula (Schleppenbach et al., 2007). A more recent set of national recommendations for best practices in middle- and secondary-school mathematics released by the What Works

Clearinghouse endorses the use of worked examples (Star, Caronongan, et al., 2015). Those recommendations specifically highlight the efficacy of examining common errors in both an individual and teacher-directed fashion.

A newer strand of investigation has extended the use of incorrect worked examples in typical short-term or lab experiments to Algebra classrooms across an entire instructional unit or entire course (Booth, 2011; Booth and Oyer et al., 2015; Star & Pollack et al., 2015). Exposure to myriad context-specific variables that surface in the day-to-day classroom setting often sullies what were significant results obtained in the laboratory. Nevertheless, it is crucial to extend research to applied situations to provide useful strategies and meaningful results to practitioners (Koedinger, Booth, & Klahr, 2013). Specific to the current study, the worked example effect has only received minimal exposure to the real-world classroom (Booth and Oyer et al., 2015).

Indeed, instruction that focuses on and includes discussion of common errors can help learners strengthen their prior knowledge by engaging both correct and incorrect mechanisms (Schleppenbach, Flevares, Sims, & Perry, 2007). According to this perspective, students do not un-learn errors but rather hone their prior knowledge to pinpoint interactions between actions and features that produced the errors (Ohlsson, 1996b). To this end, recent work has embraced errors as a means of better understanding the roots of student misconceptions within the various content domains of algebra (Booth, Barbieri, Eyer, & Pare-Blagoev, 2014; Booth and Cooper et al., 2015).

A pedagogical approach along these lines involves student anticipation of potential errors. In these exercises, students analyze unsolved problems to identify the potential errors that they or another student might make in solving them. As such, they are similar to incorrect worked examples in that they engage learners in error reflection; however, they attempt to do so in a more open-ended fashion in line with current student-centric instructional trends. Booth, Begolli, and McCann (2016) have demonstrated a connection between incorrect worked examples and error anticipation where students who reflect on errors presented in incorrect worked examples are more likely to anticipate errors another student might make. In their work, the authors suggest error anticipation as a potential means of making errors more prominent in a classroom for those teachers who lack the time or access to error-centered lessons but cautions that the underlying mechanism for growth from these exercises is unclear.

This study frames the mechanism for growth from error anticipation exercises via the lens of cognitive load theory (Sweller, 2011). Through this lens, traditional worked examples have proven to be an effective way to scaffold instruction. The goal is to minimize any negative effect on working memory imposed by the instructional materials, known as extraneous cognitive load, which interferes with schema acquisition (Sweller, 2010). However, as the learner becomes more proficient in the material worked examples encounter an expertise reversal effect where the marginal benefit of their use decreases (Kalyuga, Ayres, Chandler, & Sweller, 2003; Sweller, 2011).

Asking learners to self-explain worked examples has proven to be an effective extension of traditional worked examples for furthering schema development in the face of expertise reversal (Paas, Renkl, & Sweller, 2003). Learners with inadequately formed schema within a specific domain, known as novices, who increase their understanding and strengthen schema to become experts require instructional activities with decreasing levels of guidance to maximize learning. Self-explanation of worked examples is effective for students with both high and low prior knowledge, and the examples presented can transition from correct examples to a mixture of correct and incorrect examples to facilitate deeper schema development (Grobe & Renkl, 2007; Booth, Lange, Koedinger, & Newton, 2013). However, self-explanation of worked examples itself is susceptible to expertise reversal (Booth and Cooper et al., 2015; Booth and Oyer et al., 2015).

Minimizing extraneous cognitive load facilitates engagement in schema construction and automation, known as germane cognitive load, whereby working memory can focus on the interacting elements intrinsic to a task or concept (Sweller, 2010). Despite the initial increase in germane learning activities, both correct and incorrect worked examples will still see their effectiveness wane as learners move from novice to expert. Recent work has demonstrated that, after initial schema development through high-guidance instructional activities, learners are able to better retain subsequent information they generate themselves, though this research centers on generating correct solutions (Slamecka & Graf, 1978; Reifer, Chen, & Reimer, 2007). Error anticipation exercises that prompt students to generate possible errors a student might make in solving an equation could potentially help maximize germane cognitive load as worked examples' effectiveness fades.

The intervention of whole-class error anticipation exercises following teacher-led, high guidance correct worked examples across an instructional unit is a worthwhile avenue to explore. To account for the range of abilities in real-world classrooms, teachers need to include enough correct worked examples to ensure those with low prior knowledge are provided the opportunity to construct the necessary schema to move beyond them. Fortunately, prior work has shown that the gap between novices and experts can be eliminated through exposure to high-guidance activities, with no consequence to high prior knowledge students, in as little as two class meetings (Pollock, Chandler, & Sweller, 2002). The structure of this study was informed by these insights to include multiple teacher-led correct worked examples in every lesson across all conditions.

However, making a clear distinction between novices and experts is impractical in real world classrooms, and prior studies have used expertise reversal as a justification for moving away from correct worked examples without accounting for differences in ability (Renkl, Atkinson, & Grobe, 2004; Salden, Aleven, Schwonke, & Renkl, 2008). Practitioners typically do not have control over the makeup of their classes. Further, it is time-consuming and challenging to clearly identify and differentiate instruction for differences in ability levels. If the intent of this study is to engage in practical research in real classrooms that provides meaningful pedagogical suggestions (Koedinger, Booth, & Klahr, 2013), the results need to culminate in practical suggestions for teachers.

This study suggests that error anticipation exercises assist practitioners in transitioning from high-guidance, teacher-centered correct worked examples to more open-ended activities that should improve student outcomes. This is due to the generation effect (Slamecka & Graf, 1978; Chen, Kalyuga, and Sweller, 2015) which holds that learners who have acquired knowledge through example study find further benefits from generating their own responses to open-ended scenarios. These scenarios tend to be independent practice problems. Incorporating errors into open-ended generation exercises builds upon prior research that suggests worked example study benefits from the same (Grobe & Renkl, 2007).

Research Questions

As mentioned above, this study seeks to build upon previous studies of incorrect worked examples and explore the efficacy of error anticipation exercises, a more student-centered activity in line with reform-based pedagogy. Specifically, McCann and Booth (2014) demonstrated that students are capable of generating reasonable potential errors, but they did not explore the effectiveness of these exercises as an intervention. Subsequent work by Booth et al., (2016) suggested that teachers who would like to give errors a larger role in their classrooms, but may have limited access to error-centric lessons, could engage learners in thinking about potential errors and reflecting on what makes those errors troublesome. The current study builds upon this suggestion by implementing error anticipation as an instructional intervention where the teacher leads the class in error contemplation and reflection. The study compares this approach to correct and incorrect worked examples across multiple measures in answering the following research questions:

Research question 1: Does including errors affect student acquisition of conceptual and procedural knowledge in the Algebra domain of equation solving versus instruction exclusively with correct worked examples?

Sub-question a: Do error anticipation exercises affect student acquisition of conceptual and procedural knowledge in the Algebra domain of equation solving versus instruction with incorrect or correct worked examples.

Sub-question b: Do error-centric interventions such as incorrect worked examples and error anticipation exercises affect student acquisition of conceptual and procedural knowledge in the Algebra domain of equation solving versus instruction exclusively with correct worked examples?

Research question 2: Does including errors affect students' abilities to encode and replicate equation problem structures versus traditional instruction exclusively with worked examples?

Sub-question a: Do error anticipation exercises affect students' abilities to encode and replicate equation problem structures versus instruction with incorrect or correct worked examples?

Sub-question b: Do error-centric interventions such as incorrect worked examples and error anticipation exercises affect students' abilities to encode and replicate equation problem structures versus traditional instruction exclusively with worked examples?

Research question 3: Does including errors affect the ability and type of error predicted relative to instruction with correct worked examples?

Sub-question a: Do error anticipation exercises affect the ability and type of error predicted relative to instruction with incorrect or correct worked examples?

Sub-question b: Do error-centric instructional interventions such as error anticipation exercises or incorrect worked examples affect the ability and type of error predicted relative to instruction with correct worked examples?

Definition of Terms

- *Conceptual knowledge* refers to mathematics knowledge and facts and their interrelationships including such topics as like/unlike terms, the equals sign and equivalence, positive/negative numbers, etc. (Hiebert & Wearne, 1986; Rittle-Johnson & Alibali, 1999; Booth & Koedinger, 2008).
- *Encoding and replicating problem structures* refers to students' ability to memorize and quickly reconstruct an equation that has been displayed for a short period of time (approximately 5 seconds).

- *Error anticipation exercises* are instructional interventions utilized in this study that present the learner with a problem and asks for potential errors another student might make in solving or evaluating.
- *Extraneous cognitive load* is working memory load imposed by instructional activities that detract from learners' efficient schema acquisition (Sweller, 2010).
- *Germane cognitive load* refers to working memory load that engaged during the construction of schemas necessary for learning (Sweller, 2010).
- *Intrinsic cognitive load* is working memory load imposed by the inherent complexity of the material studied by a learner (Sweller, 2010).
- *Long-term memory* refers to the collection of schemas that are drawn upon by working memory to process information and engage in procedures (Sweller, 2011).
- *Procedural knowledge in algebra* refers to knowledge of procedures and algorithms required to engage in algebraic operations, simplifications, and equation solving (Booth & Koedinger, 2008).
- *Self-explanation of worked examples* refers to a learner's explanations utilized in examining and reconciling information contained within worked examples (Chi et al., 1989).
- *Worked examples* are example exercises that display the steps encountered in reaching their solutions. This presents the learner with the opportunity to inspect the process and see the finished product, as opposed to following written instructions.
- *Working memory* refers to the limited cognitive attention a learner can utilize to make sense of new information (Sweller, 2011).

CHAPTER 2

LITERATURE REVIEW

In this chapter, I analyze the literature related to the use of error anticipation exercises in the classroom. The chapter begins by defining cognitive load theory, the theoretical framework embraced by the study. Student outcomes in this study are measured in terms of both conceptual and procedural knowledge, so a description of conceptual and procedural knowledge in Algebra follows. I then examine the utility of errors in mathematics instruction both generally and as they pertain to self-explanation of worked examples. Finally, I will identify the gaps in the research suggested by the expertise reversal and generation effects in making a case for the inclusion of error anticipation exercises as an instructional tool.

Cognitive Load Theory

Cognitive load theory is useful for understanding the potential role of error anticipation exercises as an instructional intervention. This information processing theory utilizes human cognitive architecture to describe the ways in which learners encounter novel situations and develop approaches to reach specified goals (Sweller, 2011). It encourages instructional designers and practitioners to consider the constraints of learners' limited working memory to best facilitate the transmission of knowledge to unlimited long-term memory (Kirschner, 2002; Sweller, 1994; 2011).

Schema Acquisition

According to Sweller (1994), the two main mechanisms of learning in cognitive load theory are schema acquisition and automaticity of learned procedures. The theory posits that learners assimilate novel information, known as elements, into schemas (Sweller, 1994). These schemas represent categories of prior knowledge that closely resemble, but do not necessarily match, the new elements. The transaction is bidirectional in that new information is transformed as it is combined with information in the existing schema, and the schema itself is transformed by the newly-added knowledge. Within such a dynamic state of information, no element or schema is necessarily the same across individuals (Sweller, 2011).

For mathematics in particular, a schema can incorporate both concepts and procedures and evolves as the learner engages it over time. Although concepts and procedures are generally considered distinct types of knowledge, research supports the notion that the two exist on a continuum with some level of interaction between them (Rittle-Johnson & Alibali, 1999). As a learner progresses to more complex material, what might have been a schema composed of less-complex elements could itself evolve into an element of a more-complex schema. For example, an Algebra I student might struggle to understand how to factor quadratic functions and require practice to strengthen that particular schema. As the student becomes more proficient, he or she can progress to more advanced operations such as simplifying rational expressions and operations with rational expressions that require automaticity in factoring. This compaction of simpler schema into more-complex schema is a key process in cognitive load theory whereby less attention in the form of working memory needs to be paid to simple information or operations, freeing up the learner to concentrate on advanced topics (Sweller, 1994; 2011).

Working and Long-term Memory

Any new information must first be processed through this working, short-term memory. According to Kirschner (2002), working memory is used to process all sentient activities and is the only form of memory that learners can actively engage and monitor. It serves as a channel connecting external stimuli to long-term memory (Sweller, 2011). When encountering new information, working memory actively compares and contrasts it with any existing knowledge structures (Kirschner, 2002). Conscious attention to new information elements is necessary in order to make sense of them, identify related schema, and assimilate the information into the schemas that most closely resembles it. This is a gradual process that occurs with time as the learner is exposed to the elements and utilizes the schema (Sweller, 1994).

Throughout the knowledge acquisition process, the limited amount of working memory a learner can utilize to receive and process new information is a constraint. Previous studies have estimated the capacity of working memory at seven pieces of new information (Miller, 1956) with duration of approximately 20 seconds (Peterson & Peterson, 1959; Sweller, 2011). Schema acquisition and automation are therefore critical to minimizing the effect of such a tight window of attention (Sweller, 1994).

As the process becomes automated, the learner requires less and less deliberate attention to make sense of and utilize the information (Sweller, 1994). After much practice, the learner can apply the information with little to no working memory. Upon reaching a point of automaticity, the schema transitions from working to long-term memory, allowing for quick retrieval as necessary while at the same time freeing up working memory to extend understanding (Sweller, 1994; 2011).

Expertise

As mentioned previously, cognitive load theory submits that this acquisition and automation of vast amounts of schemas is a principal characteristic of intellectual development (Sweller, 1994). Experts, or those who have acquired and activated a large number of schema within a particular knowledge domain, have been shown to approach problems more efficiently than novices who have only recently been exposed to the information (Chi, Glaser, & Rees, 1982). For intellectually complex disciplines, experts have been estimated to have upwards of

10,000 unique schemas from which they can draw to attack a problem (Simon & Gilmartin, 1973; Sweller, 1994).

When confronted with a problem, both novices and experts examine the key surface structures for clues on how to categorize problems according to their existing schema. Novices, lacking in schema relative to experts, tend to categorize problems based entirely on those object-specific surface structures (Chi et al., 1982). To solve, novices employ a means-end strategy where differences between the problem state and the goal state are reconciled using problem-solving operators to progress backwards through sub-goals (Sweller, 1994). This stresses working memory as the learner engages a number of interacting elements and inhibits learning (Sweller, 2011).

On the other hand, experts draw upon their more-developed principle schemas composed of the lesser object schemas relied upon by novices (Chi et al., 1982). The expert does not need to examine the surface structures to categorize the problem. Any problem can be quickly matched with a stored problem configuration schema (Sweller, 1988), enabling the expert to quickly undertake the appropriate identified solution path. Bypassing surface structure characterization allows the expert to minimize or eliminate working memory utilization, freeing it up for development of deeper schemata as problem complexity progresses (Chi et al., 1982; Sweller, 1994). Within the algebra research, Star (2005) describes how students develop flexibility in solving equations where subtle problem characteristics provide cues for the most appropriate solution path. This flexibility could represent a level of automaticity where working memory use is efficiently managed via deliberate attention to only those problem characteristics.

Although the literature differentiates between experts and novices, the measuring stick that determines whether a learner is an expert, novice, or somewhere on the continuum between the two, is unclear and typically specific to the study participants. Regardless, Pollock et al. (2002) showed that novices could acquire knowledge and create requisite schema relatively quickly. They demonstrated that learners with low prior knowledge who first engaged in activities with low element interactivity, where a small number of elements are processed in working memory (Sweller, 2010), were able to gravitate to tasks that are more complex within 1-2 class meetings. So although the distinction between novices and experts is important in cognitive load theory, individuals are able to make progress toward expertise with specific tasks within a relatively short period of time, with appropriate instructional design. Particular to this study, the findings from Pollock et al. (2002) show that it is conceivable that students who are learning one particular concept or skill, for instance solving an equation that requires using the distributive property, could become proficient within that isolated skill over a class period or two.

Types of Cognitive Load

According to Sweller (2011), the two primary sources of instructional cognitive load are intrinsic and extraneous. Effective instruction minimizes extraneous cognitive load and maximizes learner engagement in the germane cognitive load responsible for schema construction (Sweller, van Merrienboer, & Paas, 1998; Paas & van Gog, 2006; Sweller 2010).

Intrinsic cognitive load represents the intrinsic complexity of the material that is essentially fixed, and cannot be altered unless the information itself is transformed (Sweller, 2011). The complexity of the knowledge being acquired is considered separate from the means in which it is acquired. The intrinsic cognitive load for a particular piece of information or task can be estimated by considering the amount of elements that need to be processed simultaneously in working memory. According to Sweller (2010), information with low element interactivity can be learned through referencing a relatively small subset of elements, thereby imposing a small strain on working memory. On the other hand, information with high element interactivity utilizes a large number of elements with many more interactions resulting in a heavy working memory load.

In beginning algebra, an example of a concept with low element interactivity could be solving a two-step equation. There may be constants, variables, operators, and equals signs, but the general form will not vary much from equation to equation. Without much practice, a learner should be able to quickly identify the important pieces that activate appropriate schema with minimal working memory. Alternately, a concept with high element interactivity could be solving a multi-step equation with variables and simplification on both sides. The learner will need to simultaneously process the varying pieces to determine a start point and progress through the steps appropriately.

The second component of working memory load, extraneous cognitive load, concerns itself with the means employed by practitioners in teaching new material. Whereas the intrinsic cognitive load is considered a fixed constant for a learner, extraneous cognitive load is a function of these instructional design factors (Sweller, 2010). The goal of effective instructional design should be the minimization of extraneous cognitive load to allow working memory to process intrinsic knowledge elements (Sweller, 2011). For example, a teacher might provide partial notes so that students only need to write the critical information. Sweller et al. (1998) go so far as to describe extraneous cognitive load as the effort a learner requires to make sense of instruction that is poorly designed.

Intrinsic and extraneous cognitive load can both trace their origin to degrees of element interactivity (Sweller, 2010). It is important to note that the interaction between intrinsic and extraneous cognitive load is best described as synergistic versus differential, and the degree of each type of cognitive load contained within a task depends on the both the instructional goals and the prior knowledge of the learner (Beckmann, 2010).

For example, in beginning algebra solving an equation with simplification and variables on both sides could impose both a high and low cognitive load depending on their background procedural knowledge as well as the instructional goals. If the goal is to teach students how to solve an equation with a variable on each side, the inclusion of additional steps that require conceptual knowledge such as using the distributive property and/or combining like terms distracts working memory from the key elements in simpler equations and imposes a heavy extraneous cognitive load on the learners. On the other hand, if the goal is to solve multi-step equations along a wide range of iterations, then the procedure would likely not impose a heavy extraneous cognitive load.

In both cases, the working memory load engaged by the learners will likely be high. If the objective is to help the students learn to solve equations with multiple steps in a variety of forms, the lion's share of cognitive load would be intrinsic. If the goal is to have students solve simpler equations, then a bigger proportion of the cognitive load is imposed by the instructional practice. According to Sweller (2011), procedures can be transformed to simplify the examples used to teach the topic so long as the transformation does not alter the intended instructional goal. Minimizing extraneous cognitive load frees up working memory to engage in schema acquisition and automation, which are the primary processes of learning in cognitive load theory. Freeing up working memory to allow the learner to engage in germane learning endeavors is the goal of instructional design within this framework (Sweller et al., 1998). It is not enough to simply decrease extraneous cognitive load. To increase germane cognitive load leaners' attention must be directed toward processes that facilitate schema construction (Sweller, 2010).

It can be difficult to design mathematics instruction that maximizes germane cognitive load for all learners in a particular class. Sweller (1994; 2010) points out that no schema is necessarily uniform across learners. Every student's schemas become particularly complex as they absorb more elements. What might be a schema to one student would be an element to a more-advanced cohort. However, this uniqueness should not preclude teachers from making instructional decisions based on their estimate of students' current schema development (Sweller, 1994).

Indeed, students' prior knowledge remains a constraint for instructional designers as demonstrated by Ayers (2006). He examined means of decreasing cognitive load by manipulating intrinsic cognitive load of a mathematics task in an Algebra 1 classroom. The study found a difference in germane cognitive load depending on the level of student prior knowledge. Reducing the complexity of the exercise resulted in students with low prior knowledge committing less errors and reporting a lower cognitive load. Students with higher prior knowledge also reported lower cognitive load for the reduced-complexity exercises but committed more errors than those who were assigned more complex tasks. Ayers (2006) opined that when the task was too simple, students with high prior knowledge were not incurring those processes of schema acquisition and automation that are germane to learning. Classroom teachers need to be aware of how the flow and structure of their lessons affect germane cognitive load for all learners. To meet all learners, simpler tasks could be deployed earlier in a lesson to accommodate those with low prior knowledge after which more-complex tasks could be introduced to facilitate germane cognitive activities in those who already had high prior knowledge.

It is therefore crucial for learning to ensure that tasks are appropriately complex. Without this nuance cognitive load theory might be incorrectly interpreted to suggest that all cognitive load must be minimized to maximize instructional efficacy. When students are appropriately engaged, cognitive load is largely directed toward the processes this theory holds responsible for learning.

Conceptual and Procedural Knowledge

The prominent role of conceptual and procedural knowledge in Algebra merits its discussion in framing the research and influencing the composition of this study. Although these types of knowledge have qualities that distinguish one from the other, the research suggests that the two work in tandem without one dominating the other although gains in conceptual knowledge tend to be held in higher regard (Hiebert & Wearne, 1986; Rittle-Johnson & Alibali, 1999). Through the lens of cognitive load theory schemas are composed of concepts and procedures. The number and complexity of the elements that compose a schema account for its effect on working memory, without explicitly labeling the type of schema as conceptual or procedural (Sweller, 2011).

Hiebert and Wearne (1986) describe mathematical competence as the ability to make connections between conceptual and procedural knowledge. Conceptual knowledge has been identified as mathematics properties and facts as well as the interactions between those bits of information (Hiebert & Wearne, 1986; Rittle-Johnson & Alibali, 1999). Within algebra, conceptual knowledge involves understanding key features such as like and unlike terms, the concept of equivalence denoted by the equals sign, and positive/negative number operations and relations, among others (Booth & Koedinger, 2008). Additionally, by this definition, conceptual knowledge includes not only understanding these pillars of content and their interrelationships, but also justifying the relevant procedures using those underlying properties and relations (Booth, 2011).

More recent work by Booth and Davenport (2013) directly examined the connections between students' conceptual knowledge and their understanding of and ability to encode problem features. Research into the importance of specific algebraic problem features has concentrated primarily on the importance of understanding the concept of equality and the equals sign (Kieran, 1981; Baroody & Ginsburg, 1993; McNeil & Alibali, 2004; Alibali, Knuth, Hattikudur, McNeil, & Stephens, 2007). However, additional problem features such as variables, negative terms, and like/unlike terms appear within any of a number of equation-solving exercises. (Kuchermann, 1978; Vlassis, 2004; Booth & Davenport, 2013).

Procedural knowledge is defined as understanding the behavior of mathematical symbols and the set of algorithms utilized to solve problems in mathematics (Hiebert & Wearne, 1986). It is strictly concerned with how to do the mathematics operations as opposed to why (Booth, 2011). Unlike conceptual knowledge, procedural knowledge is tied to specific types of problems and is therefore restricted in its ability to generalize to problems within similar domains (Rittle-Johnson, Siegler, & Alibali, 2001). In algebra, procedural knowledge includes inverse operations and related computations encountered in solving for a given variable.

Conceptual and procedural understandings are generally held to be distinct types of knowledge that nevertheless interact and exist with no identifiable border between them (Hiebert & Wearne, 1986). In this view, conceptual and procedural knowledge lie on a continuum with some level of interaction between the two types of knowledge throughout (Rittle-Johnson &

Alibali, 1999). Although many studies underscore the effectiveness of an instructional focus on conceptual knowledge, they frame the relationship between the two types of knowledge as mutually beneficial to varying degrees. Skemp (1978) defined relational understanding as knowledge of both what to do and why it works, which closely resembles conceptual knowledge.

Though the relationship is considered bi-directional, there is disagreement over the relative impact of one type of knowledge on the other, as well the specific effect of each type of knowledge on overall algebra performance. Pesek and Kirschner (2000) found that a focus strictly on relational understanding had better learning outcomes than a balance of rote procedures and relations. Further, they held that any instruction centered on instrumental understanding, defined as “rules without reasons” (Skemp, 1978), poses a danger to subsequent learning of more advanced tasks.

One school of thought holds that a gain in conceptual understanding has a greater effect on procedural knowledge than the converse. This stance dominates current mathematics education reform (Star, 2005). Rittle-Johnson & Alibali (1999) note that increasing conceptual knowledge leads to gains in procedural skill and an instructional focus on increasing conceptual knowledge leads to gains in both types. Booth and Koedinger (2008) point out that shallow understanding of problems features might convince learners to broadly apply solution structures in appropriate and inappropriate procedural situations. As such, an emphasis on the concepts underlying the approach helps learners draw deep metaphors between problem structures and appropriate procedural actions (Booth & Koedinger, 2008).

Star (2005) makes an impassioned defense of procedural knowledge in the face of what he characterizes as outright avoidance of research and discussion regarding the topic. This was not always the case, and Star (2005) points out that many mathematics educators hold procedural knowledge as valuable. These researchers argue that mathematical procedures are not simply mindless processes but complex algorithms that require heuristic decisions based on sophisticated knowledge at various points (Star, 2005; Peled & Zaslavsky, 2008). As such, procedural and conceptual knowledge are not two entirely distinct, separate forms of knowledge.

Star (2005) argues that the definitions of both conceptual and procedural knowledge have evolved over time in a way that limits procedures to algorithms whereas conceptual knowledge includes deep connections. He contends that knowledge type and quality have been conflated. Unravelling type and quality allows one to consider what constitutes deep conceptual and deep procedural knowledge and where the two may overlap. Star (2005) points out that deep procedural knowledge promotes flexibility in equation-solving that helps students interpret subtle problem characteristics.

Although Star describes flexibility in terms of finding efficient ways to correctly solve equations regardless of form, it is conceivable that understanding possible errors could bolster flexibility by directing students away from solution paths that could lead to common mistakes. For example, when solving an equation with a variable and constant on both sides (e.g., $2x - 1 = 5x - 10$), a student whose flexibility was improved via error anticipation might first consider eliminating a variable term on each side so that the result is a two-step equation, which would be easier to solve than an equation that has an isolated variable term on one side. Further, that student could might also decide to subtract the $2x$ from both sides so that the resulting coefficient of x is positive, thereby eliminating the possibility that the student would make a negative sign error in solving.

Both the above-mentioned works make sure not to downplay the importance of conceptual knowledge; their intent was only to heighten awareness of the value of procedural knowledge in the face of criticism and disregard. Though each makes a compelling case for the existence of a bias towards exclusive emphasis on conceptual knowledge, their arguments for a deeper form of procedural knowledge involves the type of connections that are a staple of conceptual knowledge. A more workable solution to this quandary is to acknowledge the interwoven nature of conceptual and procedural knowledge.

Rittle-Johnson et al. (2001) agree by framing learning as the development of conceptual and procedural knowledge in an iterative fashion. The authors determined that higher levels of initial conceptual understanding led to not only increased procedural knowledge, but also stronger performance on transfer problems. However, they also pointed out that the relationship might be bidirectional in that improving procedural knowledge contributed to gains in conceptual understanding. (Rittle-Johnson et al., 2001).

According to Rittle-Johnson et al. (2001), the catalyst for improving procedural knowledge was the level of accuracy in problem representations; good learners used their conceptual knowledge to make sense of the problem situations and further their understanding. Therefore, it is worthwhile to model and facilitate growth in representing problems to ensure meaningful progression in each student's relative area of intellectual need (Rittle-Johnson et al., 2001).

An iterative approach translates well into practitioners' classrooms. Rittle-Johnson and Koedinger (2009) found value in instruction delivered in an iterative format as opposed to a more common concepts-before-procedures arrangement. Such a format allows teachers to be responsive to students' shortcomings in either procedural or conceptual knowledge (Richland, Stigler, and Holyoak, 2012). Hiebert and Grouws (2007) noted that an iterative approach is more common in international classrooms, particularly countries whose students generally outperform those in the United States. The benefits of such an instructional sequence serves to provide a more holistic understanding of the material and better-equip students to extend and generalize their knowledge of both concepts and procedures (Rittle-Johnson & Koedinger, 2009). In this case, the instructional format fosters the development of students into the "good" learners described by Rittle-Johnson et al. (2001).

Whether implicitly or explicitly guiding learners to develop their cognitive processes, research underscores the importance of recognizing the interrelationship between conceptual and procedural knowledge (Rittle-Johnson et al., 2001; Hiebert & Grouws, 2007; Rittle-Johnson & Koedinger, 2009). This assertion is not intended to downplay the utility of measuring the effect of interventions on these types of knowledge separately, nor is it intended to dismiss the idea that one type of knowledge could be considered preeminent. Rather, the idea is to value the contributions of both types of knowledge to student learning and researchers' understanding of such an important phenomenon. To solely focus on either procedural or conceptual knowledge without any consideration of the other would paint a limited picture of student mathematics learning.

The current study makes distinctions between conceptual and procedural understanding when interpreting student outcomes. As will be described in greater detail later, the literature regarding worked examples and self-explanation often makes the same distinction between the two types of knowledge with minimal, if any, consideration for how they might overlap (Rittle-Johnson, 2006; Booth, Koedinger, & Siegler, 2007; Booth & Davenport, 2013; Boothet al. , 2013; McCann & Booth, 2014; Booth and Cooper et al., 2015). Despite the lack of consensus regarding the value of procedural versus conceptual knowledge in mathematics, it is worthwhile to understand how an instructional intervention such as error anticipation might promote one type of knowledge versus the other to further the narrative that inspired this investigation.

The Utility of Errors in Mathematics Instruction

The use of errors in the mathematics instruction has been long examined but recently celebrated. Allwood (1984) identified error detection and suspicion as an indicator of problem-solving ability. The greater a student's sensitivity to errors, the better he or she performs at solving problems irrespective of whether or not their suspicion was correct (Allwood, 1984). This highlights the notion that students' heightened scrutiny of their work will pay dividends as they repeatedly question and check calculations and procedures, offering mental repetition that fosters comfort and proficiency. According to Ohlsson (1996a), most initial knowledge of a task is general. The information gained from experiencing and analyzing errors helps to suppress the actions that create them and focus general knowledge to task-specific information (Ohlsson, 1996a; Siegler, 1996).

The value of errors in mathematics education has been examined since the 1920s, though the earliest research revolved mainly around errors in arithmetic (Radatz, 1979; 1980). According to Schleppenbach et al. (2007), the prominence of error analysis in mathematics was tempered by long-held views of early researchers such as Thorndike (1922) that errors should be extinguished. Resnick and Ford (1981), citing Thorndike (1922), further described drill and practice as a vehicle for developing automaticity and thus improving the efficiency of cognitive functioning. However, in the 1990s, reform-based curricula and initiatives began to espouse the positive virtues of embracing errors as learning opportunities (Schleppenbach et al., 2007).

Almost immediately, studies began demonstrating the value of error analysis in the mathematics classroom. Several of the more prominent works focused on inquiry learning classrooms. Borasi's (1994) seminal case study of two struggling secondary mathematics students demonstrated how student errors could be used as "springboards for inquiry". In this case, the error analysis was conducted in a cooperative fashion between teacher and student, with the resulting dialogue being the main source of data for the study. According to the study, error analysis conducted in such a manner both aided the students in learning the specific content covered in the exercises and deepened their understanding of mathematics as a whole.

Despite the exposed benefits of using errors to deepen students' understanding, teachers are still hesitant to embrace them. Silver, Ghousseini, Gosen, Charalambous, and Strawhun (2005) found that teachers avoid errors due to their potential to confuse students, particularly those who struggle mathematically. On the other hand, teachers who admit to believing in the value of errors and other instructional reforms might avoid them in practice when faced with implementing them (Leatham, 2006).

American mathematics teachers in particular struggle with embracing and using errors. Schleppenbach et al. (2007) examined the approach to errors taken by American and Chinese teachers and found that although students from both countries made errors at similar rates, Chinese teachers tended to ask more follow-up questions in an attempt to capitalize on the learning experience. Similarly, Santagata (2004) discovered that relative to American mathematics teachers, who tended to minimize any acknowledgement of errors, Italian teachers discussed errors twice as much.

This begs an exploration as to why teachers avoid engaging errors as jumping points for deeper student learning. Bray (2011) conducted a collective case study of third-grade teachers' use of and belief in error-handling exercises in class mathematics discussions. She found that teachers varied the extent to which they used errors across three dimensions: inclusion of students' flawed solutions in whole-class discussion, addressing errors to promote conceptual understanding, and mobilizing students as a community of learners around error discussion. Teachers' content knowledge and attitudes about errors influenced their employment across these dimensions.

According to Bray (2011), teachers who avoided errors were more likely to believe their weaker students would be confused by examining incorrect responses, embarrassed if their mistakes are made public, and limited in their capacity to contribute to error-centric discussions. On the other hand, those who used errors believed conceptual knowledge was more powerful and flexible than memorized procedures, and that examining errors was beneficial for all members of the class regardless of their deficits. Bray (2011) noted that when these teachers encountered errors, they were better able to draw from their knowledge base and extemporaneously respond to them in a way that highlights the central mathematical ideas of the lesson.

One limitation of Bray's (2011) work is the focus on mid-elementary mathematics. Of the four participating teachers in her study, only one described their personal school experience with mathematics as positive. It may be that secondary mathematics teachers have had a more positive experience with mathematics. Given requirements for acquiring certification, secondary mathematics teachers are also more likely to have more extensive knowledge of the content. At the same time, the use of errors in secondary mathematics classrooms as vehicles for learning has only recently attained mainstream acceptance through reform-based initiatives (Schleppenbach et al., 2007). For example, it is unclear if deeper content knowledge means secondary teachers are more comfortable highlighting errors in their classrooms or that they know more about common ones. The current study will help to fill a gap in the literature by focusing on the use of errors in secondary mathematics classrooms.

With respect to specific algebra errors, Booth et al. (2014) examined the incidence and prevalence of a variety of errors committed by students across the topics encountered in an entire year of non-honors Algebra 1 and the degree to which these errors can predict student performance on standardized end-of-year assessments. Their study included the conceptual error categories of variables, negative sign, equality/inequality, fractions, mathematical properties, and operations. Non-conceptual errors were categorized as errors in completing arithmetic. Booth et al. (2014) found that negative sign errors were the most prevalent throughout the year across all topic areas within Algebra 1 and negative sign errors committed later in the year during advanced material were predictive of lower student end-of-year assessment scores. Given the predictive nature of these errors, research examining the use of errors as learning opportunities in secondary classrooms is warranted.

Although the literature provides ample evidence that errors are a worthwhile avenue of exploration to improve student learning, they are not stand-alone instructional activities. The current study explores their efficacy as part of an instructional strategy. As will be described in the subsequent sections, both correct and incorrect worked examples provide a vehicle for errors' instructional inclusion.

Worked Examples and Self-Explanation

The use of worked examples is regarded as an effective means of facilitating the acquisition and improvement of student problem-solving skills above and beyond teacher-centered instruction and unguided independent practice (Paschler et al., 2007; Ward & Sweller, 1990). Sweller and Cooper (1985) submit that worked examples direct learners' attention towards problem structures that can be generalized. This is in contrast to traditional problem-solving exercises and their attention to specific problem goals. These focus on the end result of the problem and not the intermediate steps that contain those relevant problem structures (Sweller & Cooper, 1985).

Cognitive load theory frames this *worked example effect* in terms of element interactivity. Leaving learners to solve a problem with minimal assistance forces them to pull from a variety of schema to make sense of the problem goals and determine appropriate first steps, enacting a heavy extraneous load. Examination of worked examples reduces the number of interacting elements, thereby reducing extraneous cognitive load and freeing up working memory for more germane tasks associated with efficient learning (Sweller, 1994; Renkl, 2005; Sweller, 2011).

Not surprisingly, learners favor worked examples, particularly as compared to explicit instructions. Their preference for worked examples over written instructions has been documented for some time (Anderson, Farrell, & Sauers, 1984). Lefevre and Dixon (1986) demonstrated that learners prefer examples to such a degree that they will skip instructions if worked examples are immediately available.

Though these early studies provide information about learner preferences, they do not provide insight into means of efficiently using worked examples within the domain of conceptual and procedural knowledge in algebra. The effectiveness of worked examples depends very much on how they are presented and what is asked of the learner. Renkl, Stark, Gruber, and Mandl (1998) revealed that worked examples are much more effective when learning is guided by prompting students to engage in self-explanation as opposed to simply presenting the full solution.

Self-explanation, defined as a student's explanations used to examine and reconcile information contained within worked examples, has been identified as an effective counterpart to those examples for increasing student understanding (Chi et al., 1989). After studying sufficient worked examples to form schema, learners can strengthen those schema via self-explanation of further worked examples. This constitutes a jumping-off point to deeper engagement of working memory with germane cognitive load (Paas et al., 2003). As an instructional technique, it is highly efficient and easily implemented by teachers (Renkl et al., 1998).

Chi et al. (1989) examined how using self-explanation when reviewing worked examples can promote near- and far-transfer of problem-solving skills according to students' pre-existing problem-solving ability. The authors observed a sample of college students who had not taken a college-level physics class study worked-out physics problems. They differentiated between "good" and "bad" problem-solvers and found distinct differences in how they engaged in disseminating the information contained within the examples. The good problem-solvers accurately monitored comprehension failures and used examples as a reference, while bad problem-solvers were concerned only with finding a means to reach a solution (Chi et al., 1989).

The focus on study techniques with strictly college-level students left room for investigation of both self-explanation's effectiveness as a learning strategy for younger students and possible means of improving "bad" problem-solvers' techniques.   Chi, de Leeuw, Chiu, and LaVancher (1994) conducted a subsequent study that addressed some of these concerns by targeting eighth-grade science students while using prompting to elicit self-explanations and measuring gains in student understanding in prompted versus unprompted participants. They found that students who were prompted to self-explain after reading outperformed those who were instructed to simply read the passage. Further, the high self-explainers who produced a large number of explanations outperformed low self-explainers.

Though Chi et al. (1994) focused on self-explanation from declarative text as opposed to worked examples, the results are meaningful to this study in that they showed those students prompted to self-explain had larger gains in understanding. The authors also examined differences in outcomes amongst prompted students by again differentiating between the students' level of self-explanation, in this case labeling them "low" and "high" explainers. They found the "high" explainers outperformed the "low" explainers (Chi et al., 1994). Such a finding echoes the recommendation of Pashler et al. (2007) to engage students in deep questioning that guide them to causal and explanatory pathways. Booth et al. (2013) found that students gained greater conceptual understanding when asked to self-explain worked examples relative to guided practice alone. It is therefore important to lay the foundation for students to improve their comfort and ability to engage in a self-explanation exercise.

Research emphasizes the effectiveness of using self-explanation and worked examples to maximize student understanding; however, there are conflicting opinions in several areas. One such area is the degree of variability in the types of worked examples examined. Low-variability problem sets present the same general problem with different numerical values, whereas high-variability differ in numbers and problem format (Paas & Van Merrienboer, 1994).

A study by Catrambone and Holyoak (1990) showed that focused worked examples targeting specific subgoals translated to nonisomorphic problems. Alternately, Paas and Van Merrienboer (1994) found that students who analyzed highly varied worked examples were able to solve subsequent problems with less effort and greater accuracy than those who either studied less-variable worked examples or engaged in a more conventional practice-problem approach. Further, their study found greater variability resulted in better transfer performance versus low-variability practice.

Another area of conflict is the resulting degree of near and far transfer that worked examples paired with self-explanation yields. Haskell (2001) defines near transfer as application of newly learned knowledge to similar, but not identical, situations. Far transfer involves the application of newly learned knowledge to very divergent situations similar to analogical reasoning (Haskell, 2001). VanLehn (1996) found worked examples to be effective for learning single principles used in near-transfer problems but suggested any attempts for generalizations should be accompanied by activities and instruction aimed at boosting students' inferential aptitude. On the other hand, Renkl et al. (1998) found both near- and far-transfer effects for learners with varying levels of prior knowledge. This effect depended upon whether the self-explanation was elicited or spontaneous, with elicitation displaying a significant positive effect for low-prior-knowledge learners in terms of near transfer while having no effect for the high-prior-knowledge group.

A third area of disagreement is in the degree and method of elicitation. Rittle-Johnson (2006) found that simply prompting a student to self-explain a problem improved both learning and transfer. On the other hand, Renkl (1999) noted that self-explanation left unchecked can create obstacles to further learning that include learners' construction of incorrect or incomplete knowledge, comprehension impasses, and overestimation of competency. To mitigate these risks, Renkl (1999) suggested introducing well-timed instructional interventions that provide enough support and feedback to facilitate learning progressions without jeopardizing the independent nature of self-explanation activities.

As previously noted, Renkl et al. (1998) utilized such interventions in a two-step process. Within the elicited self-explanation condition, subjects were initially provided a model of self-explanation with explicit hints. The subjects then examined a worked example with coaching from the experimenter. Through the coaching, the subjects were notified of any omissions and received direct assistance and feedback (Renkl et al., 1998). When used appropriately, interventions help learners operate more efficiently and with fewer errors (Catrambone and Yuasa, 2006).

Gerjets, Schieter, and Catrambone (2006) suggest reworking these examples from a conventional “molar” model that emphasizes problem categories and associated solution methods to a “modular” model that breaks down complex examples to smaller, more manageable elements that are easier to generalize back into larger problems. Essentially, a holistic, top-down approach to worked examples that focuses on sets of larger, more thorough problems and the schematic themes puts too great a burden on learners’ cognitive loads. The authors use a problem that asks participants to compute the probability of a complex event as an example. A molar approach would insist the learner to use the formula for permutations without replacement. A modular approach would guide the learners through the construction of individual event probabilities culminating with the multiplication rule. Breaking down these larger examples into smaller solution components reduces cognitive load and allows better transfer of those elements to related problem types (Gerjets et al., 2006).

To summarize, there appears to be a consensus that the elicitation of self-explanations is an effective means of increasing learning. Specifically, research suggests worked examples with self-explanation improves conceptual knowledge by encouraging learners to identify gaps in their understanding and creating new knowledge from a reconciliation of related prior knowledge and new information (VanLehn, Jones, & Chi, 1991). The degree to which the gaps are filled with new information depends on the number of self-explanations produced by the students (VanLehn & Jones, 1993; Chi, 1994). Booth et al. (2007) have demonstrated that these gaps of conceptual knowledge exist with algebra and hinder students’ ability to engage in algebraic procedures. Therefore, it is important to improve students' ability to self-explain and make sense of worked examples.

Incorrect Worked Examples

The use of incorrect worked examples is a relatively new direction of worked example and self-explanation research. A study by Grobe and Renkl (2007) asked students to self-explain both correct and incorrect worked examples. They found a benefit on measures of knowledge transfer from including errors in worked examples for students with sufficient prior knowledge versus correct worked examples. For low-prior-knowledge students, the study found that correct worked examples only were most effective.

To measure how the inclusion of errors affected cognitive load, the Grobe and Renkl (2007) study asked students to verbalize their thoughts and coded them according to whether they represented intrinsic, extraneous, or germane cognitive load. Participants in the correct and incorrect worked examples condition, regardless of prior knowledge, self-reported higher germane cognitive load. The authors clarified this paradox by hypothesizing that although learners of all types engaged in germane activities as evidenced by their self-explanations, those with high prior knowledge seemed best equipped to translate those activities into transfer gains. They specifically enunciate the idea that as learners become more proficient, correct examples become redundant as per the expertise reversal effect, which is described in more detail later in this paper. They argue that the errors force the learners to switch from example-study to problem-solving mode where they reconcile the error with the correct operator.

In a pair of experiments using the Carnegie Cognitive Tutor software, Booth et al. (2013) showed that infusion of correct and incorrect worked examples and guided practice was worthwhile even within an effective curriculum. They found that students who examined either all incorrect worked examples or a balanced mixture of correct and incorrect made greater gains in conceptual knowledge than those who only examined correct worked examples. In both experiments, Booth et al. (2013) identified the examples as correct or incorrect and prompted students to identify specific justifications for their classification as such. Their study used a high level of support and structure to elicit self-explanations through drop-down menus within the Cognitive Tutor software. Specifically, students who were examining incorrect worked examples were provided drop-down boxes from which they would choose one of several provided potential missteps. When students worked with correct examples, they were still prompted to justify what made the work correct. These researchers note that such scaffolding for explanations likely made the exercise more accessible to students regardless of ability level. There may therefore be an optimum level of support that provides an appropriate degree of scaffolding that minimizes extraneous cognitive load. Within the context of cognitive load theory scaffolds reduce extrinsic cognitive load, freeing up working memory for germane learning activities (Sweller, 2010).

A robust study by Booth and Oyer et al. (2015) examined both the scalability of worked correct and incorrect examples to larger populations and a wider range of topic areas, as well as their effect on students of varying levels of prior knowledge. The first of two experiments supplemented an algebra unit on equations with worked examples for the treatment group and found a significant effect of condition. Specifically, the study used random assignment of individual students (N = 56) split approximately evenly between two conditions that received traditional instruction but differed in their independent practice assignments. Students in the treatment group received 10-12 problems that included prompts to explain correct and incorrect worked examples as well as problems to solve. Students in the control group were asked to solve a similar number of problems. At posttest, students in the treatment condition outscored those in the control condition (Booth and Oyer et al., 2015). A second experiment scaled the experiment from three classrooms to nearly thirty and included worked examples for content that ranged from pre-algebra to quadratics. This investigation found that students with low prior knowledge seemed to benefit more from the treatment condition than those with stronger prior knowledge. However, the study did not find a significant overall effect of condition, which the authors believe may be due to the effect of worked examples varying by topic area (Booth and Oyer et al., 2015).

The current study hypothesizes that anticipating potential errors could be a logical next step beyond incorrect worked examples for increasing automaticity. When common tripping points are incorporated into schemas, the working memory that would otherwise be expended on avoiding those errors can be freed up to engage in germane cognitive tasks.

Gaps in the Literature

This study explores the effectiveness of error anticipation exercises as an instructional intervention. McCann and Booth (2014) demonstrated that students are capable of examining a problem and suggesting potential errors one might make in solving or simplifying. The investigation of the effectiveness of error anticipation as an intervention makes sense given the embrace of errors as vehicles of learning. Error anticipation exercises use a more open-ended format than worked examples but can be easily deployed as a tool to accompany more traditional types of instructional activities, thereby increasing the potential for adoption by teachers who may be wary to cede control of activities to the degree required by other student-centered practices. Further, error anticipation exercises provide an extension activity when returns on correct and incorrect worked examples begin to diminish.

The current study adopts cognitive load theory because it enunciates several effects whose influence on knowledge acquisition helps make the case for error anticipation exercises as a classroom intervention. A number of studies examining self-explanation of worked examples have used the theoretical framework overlapping waves theory, and proponents of both have engaged in a spirited, public debate in recent years (Kirschner, Sweller, & Clark, 2006; Hmelo-Silver, Duncan, & Chinn, 2007). However, the constructivist nature of overlapping waves theory insufficiently accounts for practitioner influence on knowledge development. I believe the effects hypothesized by cognitive load theory justify exploring error anticipation exercises as a logical extension of self-explanation of incorrect worked examples. Those effects are described below.

Expertise Reversal Effect

As previously mentioned, when novices study worked examples they encounter lower extraneous cognitive load due to the reduced number of interacting elements necessary to make sense of the information (Sweller, 1994; Renkl, 2005; Sweller, 2011). As the learner incorporates these isolated elements into schemas that are themselves treated as elements in more-complex schemas, the supports in place to help novices understand the information become redundant (Sweller, 2011). When expertise increases, high guidance instructional techniques experience diminishing effectiveness in an expertise reversal effect.

Continued exposure to high guidance activities could go so far as to have a negative effect on learning, as redundant supports could impose an extraneous cognitive load for experts (Kalyuga et al., 2003). To counter the expertise reversal effect, practitioners need to design instruction to wean learners off those supports. After studying worked examples, Sweller (2011) suggests a learner should transition to exercises that are more open-ended.

Specific to this study, Renkl et al. (2004) demonstrated that fading steps from worked examples and requiring learners to reconcile and produce the missing steps improved near and far transfer. Their work informs the current study in two ways. For one, the interventions were carried out over 60-75 minutes. This demonstrates that although it is implausible that students would operates as experts, in the common understanding of the term, over such a short period of time they could conceivably acquire enough knowledge that would render correct worked examples less effective. Second, the study did not consider or differentiate for ability level in the intervention or analysis. It looked at the effect of fading on all learners, which mirrors the instructional decisions for classroom teachers.

Generation Effect

According to the generation effect, learners benefit more from generating their own responses to problems than from reading correct responses to questions (Slamecka & Graf, 1978). Research on this effect has focused chiefly on vocabulary development though inroads have been made in determining the effect of self-generation in similarly-rote arithmetic tasks (Pesta, Sanders, & Murphy, 1999; Riefer, Chien, & Reimer, 2007) with benefits particularly apparent for students with low prior knowledge (Rittle-Johnson & Kmicikewycz, 2008). More recent work guided by neuroscience explores the generation effect as it applies to secondary-level math concepts, where an emphasis on engaging students in practice activities that force them to struggle improves learning outcomes with a reduced cognitive load on subsequent tasks (Wirebring et al., 2015; Jonsson, Kulaksiz, & Lithner, 2016).

This may appear contradictory to the worked example effect. The worked example effect is generally associated with acquiring conceptual and procedural knowledge, for instance in algebraic equation-solving, whereas the generation effect is associated with automaticity (Chen et al., 2015). Chen et al. (2015) demonstrated that this contradiction might be reconciled by considering the type and complexity of material used to demonstrate the worked example and generation effects. In fact, this relationship between the worked example effect and generation effect could be framed as a continuum that mirrors the expertise reversal effect. Complex material with high element interactivity is best learned with worked examples, whose scaffolding assists in reducing working memory load. Generation exercises, on the other hand, are more effective with low-element interactivity material (Chen et al., 2015).

When considering the level of element interactivity, it is important to note that cognitive load is specific to the learner. A learner will encounter high element interactivity when working with any novel material. After repeated practice and exposure, the material will transition to long-term memory for storage and use as an element in more advanced topics. Within cognitive load theory, the goal of instructional design is to maximize germane cognitive load by minimizing extraneous cognitive load (Sweller, 2010). As demonstrated above, a practitioner can accomplish this goal by first engaging learners in high-guidance activities such as worked examples and transitioning to problems and activities where learners generate their own responses (Chen et al., 2015).

The above-referenced works by Wirebring et al. (2015) and Jonsson et al. (2016) reinforce the importance of prior knowledge in the implementation of generation tasks. These studies used student populations who were approaching the end of their secondary education and had sufficient prior knowledge to engage in unguided algebra word problems requiring generation of solutions. Though neither study used the lens of cognitive load theory explicitly, they examined the effect of working memory capacity on students' ability to engage in mathematical tasks with varying levels of guidance. Both were conducted outside the classroom, and participants worked through example problems either with or without scaffolded hints depending on the condition in advance of a test administered one week later.

Both studies found that participants who engaged in low-guidance problem practice outscored those who were provided algorithmic assistance. Specifically, Wirebring et al. (2015) found that participants who solved open-ended problems exhibited more efficient brain activity germane cognitive load. Jonsson et al. (2016) noted that the open-ended iteration of the task provided effortful struggle that better facilitated knowledge transfer, much in the way that germane cognitive load is crucial for the transition from working to long term memory (Sweller, 2010).

The current study examined the efficacy of using error anticipation exercises as a form of student-generated response to a stimulus to maximize working memory in an attempt to better understand how errors can contribute to algebraic knowledge acquisition. As students become proficient in equation-solving, they experience diminishing returns from subsequent studying of worked examples. Deeper schema refinement requires exposure to a number of similar correct and incorrect problem configurations (Sweller, 2011). Error anticipation provides an opportunity to move students into generation exercises and expand exposure to a range of these configurations through consideration and reconciliation of possible errors.

As will be further described subsequently, this study examined both student growth in algebra knowledge and their ability to encode and replicate problem structures. Growth in conceptual and procedural knowledge tests whether error anticipation exercises does indeed foster germane cognitive load that leads to deeper schema development. Concurrently, development of students’ ability to encode problem structures provides insight into how error anticipation exercises engage students in schema creation to facilitate moving those learners from novice to expert (Simon & Gilmartin, 1973; Sweller, 1994).

Research Questions and Hypotheses

As previously mentioned, this study aims to address the following research questions:

Research question 1: Does including errors affect student acquisition of conceptual and procedural knowledge in the Algebra domain of equation solving versus instruction exclusively with correct worked examples?

Sub-question a: Do error anticipation exercises affect student acquisition of conceptual and procedural knowledge in the Algebra domain of equation solving versus instruction with incorrect or correct worked examples.

Sub-question b: Do error-centric interventions such as incorrect worked examples and error anticipation exercises affect student acquisition of conceptual and procedural knowledge in the Algebra domain of equation solving versus instruction exclusively with correct worked examples?

The literature demonstrates that as learners transition beyond novice status, they experience expertise reversal that necessitates a transition from high-guidance activities such as worked examples to more open-ended exercises (Sweller, 2011). One means of accomplishing this transition is to engage learners in generation activities (Slamecka & Graf, 1978). Considering errors enhances schema formation (Grobe & Renkl, 2007), and error anticipation exercises provide an opportunity to include error consideration in generation activities. As noted above, students in 9th grade Algebra, having been exposed to equation-solving in subsequent Pre-Algebra courses, should have sufficient background knowledge to facilitate the transition to low-guidance exercises over the course of the instructional unit. Guided by these insights, the following hypothesis was formed in response to this research question:

Hypothesis: Students who engage in error anticipation exercises in addition to studying worked examples will demonstrate significantly higher gains in conceptual and procedural understanding in the Algebra domain of equation-solving than those who examine correct worked examples exclusively or a combination of correct and incorrect worked examples.

Research question 2: Does including errors affect students' abilities to encode and replicate equation problem structures versus traditional instruction exclusively with worked examples?

Sub-question a: Do error anticipation exercises affect students' abilities to encode and replicate equation problem structures versus instruction with incorrect or correct worked examples?

Sub-question b: Do error-centric interventions such as incorrect worked examples and error anticipation exercises affect students' abilities to encode and replicate equation problem structures versus traditional instruction exclusively with worked examples?

Cognitive load theory submits that possessing a larger repertoire of schema available to draw upon helps learners make sense of new information (Simon & Gilmartin, 1973; Sweller, 1994). Learners with strong schema development require less working memory to process stimuli (Sweller, 1988). Low-guidance, student-generated activities such as error anticipation exercises may strengthen schema more efficiently than worked examples alone, particularly over the course of an instructional unit where repetition of high-guidance activities run the risk of expertise reversal (Sweller, 2011). Guided by these insights, the following hypothesis was formed in response to this research question:

Hypothesis: Students who engage in error anticipation exercises in addition to studying worked examples will demonstrate a significantly stronger ability to encode and replicate key equation-solving structures than those who examine correct worked examples exclusively or a combination of correct and incorrect worked examples.

Research question 3: Does including errors affect the ability and type of error predicted relative to instruction with correct worked examples?

Sub-question a: Do error anticipation exercises affect the ability and type of error predicted relative to instruction with incorrect or correct worked examples?

Sub-question b: Do error-centric instructional interventions such as error anticipation exercises or incorrect worked examples affect the ability and type of error predicted relative to instruction with correct worked examples?

As mentioned above, instructors need to consider transitioning to less-guided instructional activities such as error anticipation exercises upon learners' acquisition of schema through more-guided activities such as worked examples. Exposing students to such instruction facilitates efficient exploitation of deeper-developed schema (Grobe & Renkl, 2007). As a result, learners are better-prepared to consider and enunciate potential common errors that extend beyond simple computational mistakes.

Hypothesis: Students who engage in error anticipation exercises will produce significantly more reasonable errors that focus more directly on conceptual versus procedural errors.

CHAPTER 3

RESEARCH DESIGN AND METHODOLOGY

Context

The study was conducted at an inner-ring suburban public high school immediately outside a major city in the Mid-Atlantic region of the United States. The district has a total enrollment of approximately 4,000 students and the high school serves roughly 1,200 students. Demographically, the predominant student backgrounds are non-Hispanic Caucasian (82%), African-American (7%), and Asian (7%). The district serves a relatively affluent population, with only 15% of students classified as economically disadvantaged and a dropout rate of less than 1%. Special education students comprise 15% of the population, whereas 5% are identified as gifted.

The district has three elementary schools feeding into a single middle- and high-school. The study was conducted at the high school, which is where I am employed as a mathematics teacher. The district's instructional staff has an average of 14.8 years' teaching experience, with an average of 13.1 years' experience in the district.

The high school performs consistently well on high-stakes assessments. The percentage of students scoring advanced or proficient on state high-stakes assessments in reading and mathematics is typically at or above 80%, placing it near the top of rankings when compared to geographically proximate schools. Its academic program of study requires that all students complete three mathematics courses, though approximately 90% of the students elect to take at least four. Roughly one-third of entering ninth grade students have completed at least Algebra 1 and are placed in more advanced courses; with the exception of 10-20 students enrolled in a Mathematics Resource Room class, the remaining take Algebra 1 their ninth grade year. The identification process for placement in advanced mathematics classes begins in the late elementary grades and factors in a combination of assessment performance and teacher/parent recommendation. The student's ultimate path is determined by seventh grade, with advanced students placed in Algebra 1 while the remaining students take grade-specific mathematics courses in anticipation for Algebra 1 placement in ninth grade. There is no honors- or advanced-level Algebra 1 class offered at the high school, as students considered advanced or accelerated are presumed to have done so in the middle school.

Within these college preparatory classes, I build upon the work of McCann and Booth (2014) by examining the efficacy of using error anticipation exercises at the high school level. Their work also examined Algebra 1 classes, but those classes were at a middle school. Although there is an ongoing push to encourage students to take Algebra 1 before reaching high school (Evan et al., 2006), the vast majority of students wait until 9^{th} grade (Walston & McCarroll, 2010). As such, it is important to examine the efficacy of error anticipation exercises with the portion of students that are the last to move to advanced mathematics courses.

Participants

Class size for Algebra 1 sections at the school is capped at 25 students. A total of five sections of Algebra 1 were included in the study. All classes were rostered to capacity. Of 125 possible participants, 105 agreed to participate. All 105 participants were ninth-graders.

Students were informed that their participation was entirely voluntary. Assent forms were provided to the students who agreed to participate, and their parents were provided with consent forms. Along with study-specific measures, demographic information that included grade level and gender was collected about the participating students.

Two teachers agreed to be in the study. Teacher 1 taught two sections included in the study, both of which were included in the error anticipation condition. Teacher 2 taught three sections, including the two sections included in the incorrect worked examples condition and the single correct worked examples comparison condition. Table 1 presents their demographic information.

Table 1

Participant Demographic Information

Teacher	Section	N	Male %	Female %
Teacher 1	Error anticipation 1	25	68.0	32.0
	Error anticipation 2	25	48.0	52.0
Teacher 2	Incorrect worked examples 1	18	55.6	44.4
	Incorrect worked examples 2	18	55.6	44.4
	Correct worked examples	19	63.2	36.8
Total		105	58.1	41.9

Teachers

The total ninth-grade Algebra 1 cohort is split into 6-7 class sections depending on enrollment. The teachers' individual schedules vary according to staffing constraints. For the academic year in which this study was conducted, seven sections of Algebra 1 were split between three teachers.

Two teachers agreed to be included in the study. These teachers both had Master's degrees and at least 3 years' teaching experience at the high school and at least two years' experience teaching Algebra 1. Both teachers had used incorrect worked examples intermittently, but had never consistently used them throughout a unit.

It is important to note that there were two additional sections of Algebra that were running at the high school at the time of the study. The original intent was for those sections to be included in two of the conditions to better even out sample sizes and increase statistical power. However, the teacher assigned to those sections abruptly resigned immediately before the study was conducted. There was a rotation of temporary substitutes assigned to those sections until a long-term substitute was appointed. Due to the volatility inherent in a classroom with inconsistent instruction, those sections were excluded from the study.

The included teachers were provided consent forms and trained in both the administration of study-specific tests and measures as well as the implementation of interventions during the lessons. More specifically, I met with the teachers to adjust lessons plans to include either error anticipation exercises or incorrect worked examples depending on assignment. Appendix A through C provide a sample lesson protocol for each condition. See Appendix D for an example of the correct worked example for one section of the covered unit and the replacement incorrect worked example or error anticipation exercise.

Ethics

It is important to mention that I am currently employed as a teacher at the school in which the study was conducted. At the time of the study, I did not teach any sections of Algebra 1 but generally encounter this population of students in higher-level classes I teach. To assure students their participation was both voluntary and not subject to any punitive measures, I minimized my contact with the subjects and ensured the participating teachers had replaced any student names on study materials with unique identifiers. I had no access to any information that would allow for cross-identification. All study materials are maintained in a secure, locked box and will be destroyed after a time period of three years from the beginning of the study.

Materials

Pre- and Post-Test

Students in all conditions completed the Connected Mathematics 2: Say It With Symbols pre-test (see Appendix E) (Lappan, Fey, Fitzgerald, Friel, & Phillips, 2006). This was the same assessment used by McCann and Booth (2014) to measure students' Algebraic conceptual and procedural knowledge. The assessment includes 8 procedural items, 19 conceptual items, and two error anticipation items, and it served as both the pre- and post-test for the study.

Items were classified as procedural if the students needed only to engage in mathematical operations to answer them correctly (Booth, 2011). An example of a procedural item would be "Expand $2(x - 7)$ using the distributive property". Items classified as conceptual required students to examine the interaction of properties and facts (Hiebert & Wearne, 1986; Rittle-Johnson & Alibali, 1999). Conceptual items would include problems such as "Are $2x$ and x^2 like terms?" or "Is $x^3 = x + x + x$?", as well as asking students to consider what role each piece of $P = 20x - 500 - 7x$ might play if the equation represented profit for a fundraiser (i.e., -500 is the fixed cost to rent the room). The error anticipation item asked the students to think about the equation $5x - 2 = 8$ and name two common mistakes another student might make in solving for x.

Error Anticipation Items

The error anticipation items from the above-mentioned instrument were analyzed to determine the type of error anticipated by students in each condition at pre- and post-test. The textbook utilized by the participating school, Big Ideas Math: Algebra 1 (Larson & Boswell, 2015), includes the following topic areas in its unit on equation solving: solving simple equations, solving multi-step equations, solving equations with variables on both sides, solving absolute value equations, and rewriting equations and formulas. Within this unit, students learn to solve equations of various types for a given variable using inverse operations (e.g., solve $2x - 1 = 5$ for x). Misconceptions in solving these equations typically center on main features that include the equals sign (i.e., forgetting to balance an equation across the equals sign), variables (i.e., confusing like and unlike terms), coefficients, operations (including evaluating using order of operations and solving using inverse operations), and exponents (Booth and Koedinger, 2008).

Procedural errors include mistakes in using inverse operations, order of operations errors, and computational errors. These errors are procedural in nature due to their strict focus on mathematical operations (Hiebert & Wearne, 1986). Consider again the equation $2x - 1 = 5$. An example inverse operations mistake is subtracting 1 from each side as opposed to adding. Another such mistake would be applying different operations to both sides, for example adding 1 to one side and subtracting it from the other. For the purposes of this study, order of operations errors and inverse operation errors will be considered the same, as a two-step equation is as simplified as possible before solving. This distinction is necessary as students may posit either as an example of an anticipated error. A computational error would be a mistake in adding, subtracting, multiplying, or dividing in applying the inverse operations.

Conceptual errors in solving two-step equations center on variables and the equals sign. These errors point to deficits in conceptual knowledge due to their focus on key algebraic problem features such as like/unlike terms and the concept of equivalence (Booth & Koedinger, 2008). As previously mentioned, a variable error could include a misconception involving like and unlike terms. An example of such an error would be solving $2x - 1 = 5$ by first incorrectly combining $2x$ and -1 to create the equation $x = 5$. An equals sign error would involve applying an operation to one side of an equation only. For example, in solving $2x - 1 = 5$, an error would include adding 1 to the left side only to reach the new equation $2x = 5$.

Student-anticipated errors for these items were coded according to the themes utilized by McCann and Booth (2014) that included errors in arithmetic, order of operations, like terms, variables, negative sign, and equals sign, as well as other reasonable errors and unreasonable error/non-response. Student responses that included arithmetic and order of operations errors were classified as procedural, whereas responses that included like terms, variables, negative sign, and equals sign errors were considered conceptual. Student responses that were reasonable, but unable to be classified in any of the existing categories, were considered procedural errors for calculation purposes.

Feature encoding

Students' ability to recognize key problem features and correctly represent them, known as feature encoding, was measured using a task similar to one used in recent studies (McNeil & Alibali, 2004; Booth & Davenport, 2013). Participants in all conditions were instructed to memorize equation solving problems displayed on an electronic whiteboard. Each item was displayed for a period of 5 seconds, after which it disappeared from the screen. The participants were then asked to replicate the problem by re-writing it from memory on a sheet of paper. The participants were given as much time as needed to think about and replicate the problem, after which the next item was displayed.

The five items (Appendix F) began with a one-step equation and increased in complexity to a two-step equation, equation with distributive property, equation with variables on both sides, and an absolute value equation. The items included elements such as variables, exponents, equals sign, negative signs, parentheses, coefficients, etc. The five items contain 59 distinct features, ranging from a two-step equation with 6 elements to a multi-step absolute value equation with 14 elements. Participants completed this exercise before and after the study to assess any change in feature encoding ability.

Data analysis for encoding is informed by the work of Booth and Davenport (2013), who categorized participant encoding errors as either deep or shallow in nature. One major difference between this study and prior work with encoding is that this study used percentage correct overall and within the different classifications as the unit of measurement. This methodology was employed because of the intention to analyze improvement across pre- versus posttest performance.

Further, the equations used for the encoding exercises contained a wider range of elements such as grouping symbols, exponents, etc. This study considers coefficients/constants and variables to be shallow problem features. Positive/negative signs, equals signs, grouping symbols, and exponents are classified as deep problem features because their inclusion or omission fundamentally changes the nature of the equation. Exponents were considered deep and not merged with coefficients/constants because the beginning Algebra students in the study likely had minimal exposure to equations with exponents in their pre-Algebra experience. This study examines overall encoding performance as well as encoding performance of shallow and deep problem features.

To streamline analysis, student responses were compared to the displayed equation. Correctly reconstructed elements were counted towards the student score. For example, the first encoding item was the equation $3x - 4 = 82$. The 3, x, 4, and 82 were classified as shallow features whereas the negative and equals signs were considered deep features. In order to be scored correct, each feature needs to be in its appropriate place. For instance, if a participant moved the 4 across the equals sign it would be marked as incorrect.

Procedures

As mentioned above, the study employed five sections of Algebra 1 split between two teachers. One teacher was assigned three sections of Algebra 1, with the other teaching two sections. The study randomly assigned two classes to the treatment group and two classes to the incorrect worked examples comparison group. After consulting with building administration, the treatment and incorrect worked examples comparison sections were not split between teachers. The building principal wanted the teachers involved in the study to have consistent methodologies due to the sensitive nature of algebra 1 achievement scores. As previously noted, the teachers enrolled in the study both had Master's degrees and at least 3 years' teaching experience at the high school level. Administration encouraged a cooperative team approach to teaching Algebra 1, and both teachers had worked together closely over the previous two years to develop lessons and assessments that were nearly identical. The two teachers also kept a similar pace and were typically within 1-2 days of each other throughout the school year. Using random assignment, the teacher with two classes was assigned the error anticipation sections and the teacher with three classes was assigned the incorrect worked examples comparison group. The teacher with three sections was also assigned the correct worked examples comparison group.

The included teachers were provided consent forms and trained in both the administration of study-specific tests and measures as well as the implementation of interventions during the lessons. More specifically, I met with the teachers to adjust lessons plans to include either error anticipation exercises or incorrect worked examples depending on assignment. Appendix A through C provide a sample lesson protocol for each condition. See Appendix D for an example of the correct worked example for one section of the covered unit and the replacement incorrect worked example or error anticipation exercise. To minimize any potential teacher effect, the participating teachers were trained to ensure they understood exactly how to implement each condition. I provided scripts for the error anticipation and incorrect worked example items and observed one class of each condition to ensure teacher fidelity.

The study took place over an instructional unit on equation-solving. The study commenced on a Monday, during which time pretests were administered, and concluded with posttests on the Friday of the following week. There were 5 sections covered over the course of the study. Each section required approximately two instructional days; one day for teacher-led instruction and guided practice and one day for independent practice or inquiry-based activities. There was also approximately a half-day apiece for a small formative quiz and test review. The students took a summative classroom-based test on the Monday following the conclusion of the study.

During teacher-led instruction, all three conditions examined 3-4 worked examples per lesson. The only difference between the conditions was in the final example. The treatment group received instruction that replaced a teacher-led correct worked example with one teacher-led error anticipation exercise per lesson (see Appendix A). One comparison group received the same instruction and assignments as the treatment group, but replaced a teacher-led correct worked example with one teacher-led incorrect worked example per lesson (see Appendix B). The second comparison group received the same instruction as the above-mentioned two, but without replacing a teacher-led correct worked example (see Appendix C). Therefore, the students in both the incorrect worked examples and error anticipation sections were exposed to 5 instances of the intervention occurring once every two days.

The study compares the relative effectiveness of employing error anticipation versus incorrect worked examples as an instructional intervention. This study is unique in that it employs error anticipation as an active pedagogical technique used during instruction as opposed to previous studies that engaged students in self-explanation of incorrect worked examples or used error anticipation as an independent student exercise after instruction.

I met with the participating teachers prior to implementing the study to provide guidance regarding the appropriate structure of any intervention depending on condition. It is important to note that the participating teacher in the treatment group was instructed to provide no supports, suggestions, or prompts aside from asking students about potential errors. Doing so would too closely approach engaging in an incorrect worked examples exercise and compromise the comparison. However, given the wide range in complexity of the equations covered within the unit, there were several opportunities to prompt students for potential errors. As such, the teacher was instructed to prompt for errors, solicit student responses, correctly complete the step, and again prompt for errors before working through the next step.

This study used a pool of students whose academic performance and mathematics aptitude will be very similar due to the fact that the students have been subject to close scrutiny that resulted in their placement in a course of study leading to Algebra 1 in ninth grade. Although the study cannot speak directly to the process and deliberations between those responsible for placement, it can be reasonably assumed that the makeup of the assigned classes will be homogenous enough to warrant a quasi-experimental study.

Random assignment of students to treatment and control groups would have been ideal. Unfortunately, the constraints inherent in conducting research within a functioning school setting make such a process infeasible. Slavin (2007) points out that quasi-experimental research, where matched groups are compared in the absence of truly random assignment, is a practical work-around when faced with these restrictions. In such a design, it is imperative to ensure the treatment and comparison groups are as parallel in as many areas as possible (Slavin, 2007).

After consulting with the participating teachers and building administrators, it was decided to administer the study instruments over two consecutive days both prior to and immediately following the instructional unit in order to minimize instructional disruption. In both cases, the participants were administered the test instrument on the first day of testing followed by the encoding exercise on the second day. The participating teachers administered both instruments to the full classes. The teachers were encouraged to closely monitor student behavior during the encoding exercises to ensure students were not copying the equations while they were still displayed to ensure validity of the results.

I observed one class each for all three conditions on the Wednesday after commencement of the study. The same teacher taught the correct and incorrect examples, and I observed consecutive classes that used correct worked examples and incorrect worked examples respectively. The time structure of the class was the same for both. The teacher put a warm-up activity on the board for the students to work on while homework was checked for completion. This was followed by querying and answering any questions the class had regarding the homework as well as reviewing the warmup activity. The teacher then transitioned into correct example study. This lesson covered multi-step equations, and in both sections the teacher displayed the example on a whiteboard using an interactive web-based version of the text. As the teacher clicked through the example, each step and accompanying justification appeared. The tenor of the class during these examples was almost exclusively teacher-centered. The only question asked during any worked example in the two sections involved checking solutions;

namely, the student asked if they were to be required to check any answers on assessments in the manner that the textbook example displayed. The teacher replied that checks would be encouraged but not required.

At the third example, the classes diverged. The correct worked examples class saw the third example from the textbook. In the incorrect worked examples class, the teacher displayed a multi-step equation problem from the text materials and asked the students to find and correct the error, then solve the equation. The error in this problem occurred where $-2(7-y)$ was simplified to $-14-2y$. The teacher waited for approximately 1-2 minutes before querying the class. The first student to answer correctly identified the error, which was that -2 was incorrectly multiplied with $-y$ to get $-2y$. A few students expressed frustration that they did not notice the error. The teacher remarked that this was a common error he encountered when grading tests and quizzes and encouraged the students to pay attention whenever there's a negative term that is being distributed into parentheses. He noted that most of these errors usually involve the second term in the parentheses. Students will correctly multiply the first term by the negative but forget to include the negative in multiplying the second term. He cautioned the students to be careful when distributing, particularly when there is a negative number being distributed.

The teacher then worked through the correct solution on the electronic whiteboard. After completing the final worked example in each class, the teacher wrote the homework assignment on the board and implored the students to get started in class. The teacher spent the final 15 minutes of class circulating around the room, providing assistance, and managing student disruptions.

Later in the day I observed an error anticipation section that covered the same lesson. The lesson was almost identical in both the warmup and the first two worked examples. The only difference was in the presentation, as this teacher used prepared PowerPoint presentations to work through the same examples. For the third worked example, the teacher displayed a slide with the prompt "describe an error another student might make in solving the equation $2(1 - x) + 3 = -8$, then solve". The teacher circulated around the room for approximately 2 minutes as the students worked on their own.

When it came time to share out, the teacher stood in front of the electronic whiteboard and solicited responses from the class. The first response came from a student who suggested that another student might not distribute the 2 to both terms. The teacher agreed that this was the first error she thought of when she saw the problem. A different student suggested that another student might distribute the 2 to all of the terms of the left side of the equation. The teacher then wrote out on the electronic whiteboard what it would look like if a student distributed to all three terms. She outlined how doing so would affect the answer by artificially inflating the 3. She then solicited the class for one more error, and a student mentioned that the 8 was negative so another student might mess up when adding or subtracting from both sides. Another student commented that he/she always seems to make a mistake with something like that. The teacher used that opportunity to talk about what it means to add to or subtract from a negative number and described strategies for addressing those types of problems. The teacher then worked through the correct solution steps, after which she asked the students to grab one of the class set of textbooks and begin working on the homework assignment. She spent the remaining 8-10 minutes circulating around the room offering help on the problems.

Analysis of data

The primary method of analysis in the study was analysis of covariance (ANCOVA). According to Slavin (2007), comparing posttest scores without considering baseline scores presents the opportunity for type II errors. ANCOVA is appropriate for a pretest to posttest design as it accounts for the differences in individuals' ability in the measured variable (Slavin, 2007). Prior to ANCOVA testing, independent-samples t-tests were used to ensure individual sections within the conditions have no significant difference at pretest before collapsing into single conditions. Data analysis began with a three-condition analysis to examine the effect of error anticipation versus incorrect and correct worked examples. To contribute to the existing literature on errors in the classroom and worked examples, secondary analysis examined whether students in the combined error anticipation and incorrect worked examples conditions have significant differences at posttest relative to students who are exposed only to traditional, correct worked examples.

Unequal sample sizes

The inherent differences in sample sizes created by utilizing uneven treatment and comparison groups creates a hurdle to equivalence testing. According to Rusticus and Lovato (2014), unequal sample sizes can lead to differences in variability that translate into decreases in statistical power of results and increases in the possibility of type I error. As described in Chapter 4, all data used in means testing was first checked for ANCOVA assumptions. To counteract any violations of ANCOVA assumptions that surface during analysis, Welch ANOVA was used.

Welch ANOVA is appropriate for minimizing Type I errors whenever there is heterogeneity of variance, particularly when there are unequal sample sizes (Howell, 2002). This is a one-way ANOVA, which complicates analysis in a pre- to posttest design. In order to account for individual differences at pretest, I computed difference scores for any dependent variable that violated ANOVA assumptions and used these scores as the dependent variable in a Welch ANOVA. Although there are concerns about reliability of change scores (Oakes & Feldman 2001; Thomas & Zumbo, 2011), this study mitigates those concerns by using them only for variables that violate ANOVA assumptions and analyzing any differences using a robust test.

Every ANCOVA analysis in this study was preceded by checks for ANCOVA assumptions, with an emphasis on homogeneity of variance. Variables that did not violate assumptions were analyzed with ANCOVA. For any variable that violates assumptions, I computed a change score that was examined with Welch's ANOVA. Any significant results were analyzed with the appropriate post-hoc test.

Research Question 1

This research question was answered with an Analysis of Covariance (ANCOVA) to determine differences at posttest overall score as well as score on procedural and conceptual items. Pretest performance serves as a covariate. The analysis tested the hypothesis that students engaging in error anticipation exercises will have significantly higher posttest scores when accounting for pretest performance.

Research Question 2

This question was answered by performing Welch ANOVA comparing change scores across the conditions. Comparisons were made on overall encoding as well as encoding performance on shallow and deep problem features. As mentioned previously, coefficients/constants and variables were classified as shallow problem features whereas positive/negative signs, equals signs, grouping symbols, and exponents were considered deep problem features. This analysis tested the hypothesis that students engaging in error anticipation exercises will have significantly higher posttest encoding scores when accounting for pretest performance.

Research Question 3

This question was first approached by categorizing student responses according to error themes. As described previously, the error themes were adopted from McCann and Booth (2014) and included arithmetic, order of operations, like terms, variables, negative sign, equals sign, and other reasonable error. Student omissions or incorrect answers were coded unreasonable.

After summarizing student responses according to their error themes, differences in conditions at posttest are again investigated using both ANCOVA with pretest performance as the covariate and Welch ANOVA depending on whether the data violates ANCOVA assumptions. Analyses included differences in overall pretest scores and within procedural and conceptual errors. These analyses were conducted to test the hypothesis that students in the error anticipation condition would better anticipate errors than those in the other conditions.

Additional Analyses

The variety of student data analyzed in this study presents an opportunity to explore possible relationships between variables that have otherwise not been explored in the literature. Indeed, correlational analyses allow exploration of relationships across a multitude of variables at once to investigate potential areas of inquiry that were not explored in an experimental study (Slavin, 2007). Previous work has examined correlations between student procedural/conceptual knowledge and errors anticipated (McCann & Booth, 2014) as well as between conceptual feature knowledge, equation solving, and feature encoding (Booth & Davenport, 2013). This study capitalizes on the inclusion of pretest/posttest performance, encoding exercises with a larger repertoire of problem features, and error anticipation responses to explore new avenues of correlation.

CHAPTER 4

RESULTS

In this chapter, I present the results from the data collected over the course of the study. In the first section, I present participant demographic data and initial data analysis in advance of statistical tests. The second section presents the findings from analyses conducted to answer the three research questions. The final section presents correlational analyses that supplement the main research questions of the study. This study employs an alpha level of .05 for all significance testing.

Sample Characteristics

As mentioned in Chapter 3, the study instruments were administered over the course of two days each before and after the instructional unit. The study enrolled 105 students but due to the variability in attendance and constraints in administering make-up testing, several participants had incomplete data. Any participant who completed the before and after measures for either the paper test or the encoding exercise were included in the analysis. If a student completed one measure, but not the other, their data was included for the completed measure and excluded for the incomplete one. For correlational analysis, the only participant data included was for those who completed the pre- and posttest paper test and encoding exercises. Aggregate sample characteristics are displayed in table 2.

Table 2

Descriptive Statistics on Study Measures

Variable	*N*	Mean	*SD*	Min	Max
Pretest algebra knowledge	100	.39	.12	.13	.67
Posttest algebra knowledge	105	.40	.12	.10	.77
Pretest encoding	98	.56	.13	.14	.88
Posttest encoding	101	.63	.16	.03	.97
Pretest error anticipation	99	1.45	.76	0.00	2.00
Posttest error anticipation	99	1.37	.83	0.00	2.00

The following section includes analysis in preparation for using statistical tests in answering the research questions. Data is checked to ensure equal groups before collapsing into singular conditions, as well as for any violations of assumptions for the specific tests used in analysis.

Research Question 1

Does including errors affect student acquisition of conceptual and procedural knowledge in the Algebra domain of equation solving versus instruction exclusively with correct worked examples?

Initial Analyses

The study was conducted using five sections of ninth-grade Algebra 1. Of the 105 students enrolled in the study, 99 completed both the pretest and the posttest. Two sections each were assigned to the error anticipation and incorrect worked examples conditions, with the remaining section assigned to the correct worked examples only condition. The conditions were not split between the two teachers, so each teacher was assigned two sections of a single condition. I first tested the null hypothesis that there was no significant difference between the individual sections in the error anticipation and incorrect worked examples conditions before collapsing sections together.

For the error anticipation sections, there was no significant difference between scores on the overall pretest, t(42) = -1.045, p = .302; pretest conceptual items, t(42) = -1.483, p = .146; and pretest procedural items t(42) = -0.137, p = .892. Likewise, the incorrect worked examples sections saw no significant difference on the overall pretest, t(34) = -1.784, p = .083; pretest conceptual items, t(34) = -1.712, p = .096; and pretest procedural items, t(34) = -.427, p = .672. As a result, I accept the null hypothesis that there is no difference by section, and those sections were collapsed into single conditions for further analysis. Pretest and posttest performance by condition are displayed below in table 3.

Table 3

Pretest and posttest proportion correct (n = 99)

	Error anticipation (n = 44)		Incorrect worked examples (n = 36)		Correct worked examples (n = 19)	
Variable	Mean	SD	Mean	SD	Mean	SD
Pretest overall score	.41	.12	.34	.11	.40	.12
Posttest overall score	.41	.14	.37	.13	.40	.10
Pretest procedural score	.34	.23	.29	.17	.35	.18
Posttest procedural score	.36	.18	.31	.18	.32	.19
Pretest conceptual score	.45	.12	.36	.14	.41	.14
Posttest conceptual score	.43	.15	.40	.15	.45	.11

As described in Chapter 3, I examined the effect of exposure to errors versus traditional instruction with correct worked examples by collapsing the error anticipation and incorrect worked examples into a single, combined error condition. Before collapsing those sections, I tested the null hypothesis that there was no difference in change scores across the error conditions. Independent-sample t-tests indicated significant differences between the error anticipation and incorrect worked examples conditions at pretest for the overall pretest, $t(78) = 2.730$, $p = .008$; and pretest conceptual items, $t(78) = 2.898$, $p = .005$. As a result, I rejected the null hypothesis that there was no significant difference between the conditions at pretest. However, ANCOVA remains an appropriate tool for data analysis as the differences between the conditions are controlled for by using pretest scores as a covariate (Howell, 2002). The resulting scores for the combined error conditions on the overall pre and posttests as well as procedural and conceptual items are displayed below in table 4.

Table 4

Pretest and posttest proportion correct, combined error conditions (n = 99)

	Combined error conditions (n = 80)		Correct worked examples (n = 19)	
Variable	Mean	SD	Mean	SD
Pretest overall score	.38	.12	.40	.12
Posttest overall score	.39	.14	.40	.10
Pretest procedural score	.32	.20	.35	.18
Posttest procedural score	.34	.18	.25	.20
Pretest conceptual score	.40	.13	.41	.14
Posttest conceptual score	.42	.15	.45	.11

In advance of ANCOVA, the dependent variables were tested to ensure they did not violate the assumption of homogeneity of variance. As the tables below show, there was no significant result for either condition. As a result, I accept the null hypothesis that the variances among the conditions were equal. Table 5 displays the results of Levene's test using three conditions. Table 6 displays the results of Levene's test using a combined error condition and a correct worked example condition.

Table 5

Results of Levene's test with 3 conditions

Variable	Measure	*F*	*df*	Sig.
Overall score	Posttest	.902	2, 96	.409
Procedural items	Posttest	.067	2, 96	.935
Conceptual items	Posttest	.717	2, 96	.491

Table 6

Results of Levene's test with combined error conditions

Variable	Measure	*F*	*df*	Sig.
Overall score	Posttest	1.839	1, 97	.178
Procedural items	Posttest	.091	1, 97	.763
Conceptual items	Posttest	1.128	1, 97	.291

Data Analysis

To explore this research question, I first ran a one-way ANCOVA across the three conditions on posttest scores using pretest scores as the covariate. The results showed no significant effect of condition on posttest performance overall and within procedural and conceptual items across all three conditions. The results for the ANCOVA for each variable is displayed below in table 7. Based on these results, I accept the null hypothesis that there was no difference at posttest across the three conditions. Students in the error anticipation, incorrect worked examples, and correct worked examples scored similarly on the posttest overall as well as within the conceptual and procedural items.

Table 7

One-way ANCOVA conducted by condition for posttest scores, 3 conditions

Variable	df	F	*p*	Partial η^2
Overall score	2, 95	.025	.975	.001
Procedural items	2, 95	2.277	.108	.046
Conceptual items	2, 95	.396	.674	.008

Secondary analysis was conducted comparing the combined error conditions against the correct worked examples condition. The results are displayed below in table 8.

Table 8

One-way ANCOVA conducted by condition for posttest scores, 2 conditions

Variable	df	F	*p*	Partial η^2
Overall score	1, 96	.049	.824	.001
Procedural items	1, 96	3.630	.060	.036
Conceptual items	1, 96	.800	.373	.008

As demonstrated above in table 8, there was no significant difference between the conditions where $\alpha < .05$. However, there is a difference approaching significance for procedural items. Students in the combined error conditions tended to outscore those in the correct worked examples condition on procedural items. Post-hoc computation yields a partial η^2 of .036, which falls between a small and moderate effect size (Howell, 2002). As a result, I will accept the null hypothesis that there is no significant effect of condition for scores on overall test performance and conceptual items. However, given the result approaching significance and large effect size, I will neither accept nor reject the null hypothesis that there is no difference in procedural item scores according to condition.

Research Question 2

Does including errors affect students' abilities to encode and replicate equation problem structures versus traditional instruction exclusively with worked examples?

Initial Analyses

As mentioned in Chapter 3, this study examines participant encoding at the level of individual problem features as well as shallow versus deep problem features. Of the 105 students who agreed to be in the study, 95 completed the encoding exercises at both pretest and posttest. This study measured the number of correctly replicated problem features, which differs from prior encoding work that classified encoding errors (Booth & Davenport, 2013). To begin investigating differences by condition, I first ensured the multi-section conditions had no significant difference in encoding performance at pretest before collapsing sections into single conditions. Across the two error anticipation sections there was no significant difference in overall pretest encoding scores, $t(39) = -.839$, $p = .407$; shallow feature encoding scores, $t(39) = -1.028$, $p = .310$; and deep feature encoding, $t(39) = -.589$, $p = .559$. The two incorrect worked examples sections also had no significant difference at pretest in overall encoding, $t(33) = -1.347$, $p = .187$; shallow feature encoding, $t(33) = -.932$, $p = .358$; and deep feature encoding, $t(33) = -1.514$, $p = .140$. I accepted the null hypothesis that there was no significant difference at pretest in the sections within those conditions and collapsed those sections into single conditions. Table 9 displays performance for individual problem features and table 10 displays pretest and posttest encoding performance for the three conditions using consolidated feature types.

Table 9

Encoding performance by problem feature (3 conditions)

		Error anticipation (n = 41)		Incorrect worked examples (n = 35)		Correct worked examples (n = 19)	
Time	Problem feature	Mean	SD	Mean	SD	Mean	Sd
Pretest	Variables	.58	.19	.59	.17	.65	.23
	Coefficients/constants	.56	.11	.57	.12	.57	.15
	Signs	.49	.14	.44	.13	.51	.17
	Grouping symbols	.68	.22	.66	.22	.75	.21
	Equals sign	.60	.23	.62	.25	.82	.20
	Exponents	.32	.22	.36	.27	.33	.24
Posttest	Variables	.68	.19	.65	.19	.65	.26
	Coefficients/constants	.65	.14	.66	.14	.60	.26
	Signs	.56	.16	.57	.17	.57	.23
	Grouping symbols	.79	.23	.74	.20	.69	.30
	Equals sign	.74	.24	.72	.21	.64	.33
	Exponents	.42	.24	.41	.23	.47	.33

Table 10

Pretest and posttest encoding performance (3 conditions)

	Error anticipation (n = 41)		Incorrect worked examples (n = 35)		Correct worked examples (n = 19)	
Variable	Mean	SD	Mean	SD	Mean	SD
Pretest overall encoding	.55	.13	.55	.13	.60	.12
Posttest overall encoding	.64	.14	.64	.14	.61	.25
Pretest shallow encoding	.56	.13	.57	.12	.59	.16
Posttest shallow encoding	.66	.14	.66	.14	.61	.25
Pretest deep encoding	.53	.14	.52	.15	.60	.11
Posttest deep encoding	.63	.16	.62	.15	.60	.25

In preparation for secondary analysis comparing combined error conditions to traditional instruction conditions, the error anticipation and incorrect worked examples conditions were checked for significant differences at pretest. Independent-samples t-tests demonstrated that there was no significant difference in pretest scores, $t(74) = .098$, $p = .922$; pretest shallow feature encoding, $t(74) = -.253$, $p = .801$; and pretest deep feature encoding, $t(74) = .399$, $p = .691$. I therefore accepted the null hypothesis that there was no difference in the combined error conditions at pretest and collapsed them into a single condition. Table 11 displays pretest and posttest encoding performance by problem feature for the two conditions. Table 12 displays pretest and posttest encoding performance for consolidated feature types.

Table 11

Encoding performance by problem feature, combined error conditions

		Combined error conditions (n = 76)		Correct worked examples (n = 19)	
Time	Problem feature	Mean	SD	Mean	Sd
Pretest	Variables	.58	.18	.65	.23
	Coefficients/constants	.56	.12	.57	.15
	Signs	.46	.14	.51	.17
	Grouping symbols	.67	.22	.75	.21
	Equals sign	.61	.24	.82	.20
	Exponents	.34	.24	.33	.24
Posttest	Variables	.66	.19	.65	.26
	Coefficients/constants	.66	.14	.60	.26
	Signs	.56	.16	.57	.23
	Grouping symbols	.77	.22	.69	.30
	Equals sign	.73	.23	.64	.33
	Exponents	.42	.24	.47	.33

Table 12

Pretest and posttest encoding performance, combined error conditions

	Combined error conditions (n = 76)		Correct worked examples (n = 19)	
Variable	Mean	SD	Mean	SD
Pretest overall encoding	.55	.13	.60	.12
Posttest overall encoding	.64	.14	.61	.25
Pretest shallow encoding	.57	.12	.59	.16
Posttest shallow encoding	.66	.14	.61	.25
Pretest deep encoding	.53	.14	.60	.11
Posttest deep encoding	.62	.15	.60	.25

In advance of using ANCOVA to examine any differences by condition, the dependent variables were checked for violations of assumptions. Analysis found multiple violations of the assumption of homogeneity of variance on posttest measures. There were unequal variances across the two- and three-condition analyses on overall, shallow, and deep feature encoding. Table 13 displays the results of Levene's test using three conditions. Table 14 displays the results of Levene's test using a combined error conditions and a correct worked example condition.

Table 13

Results of Levene's test with 3 conditions

Variable	Measure	*F*	*df*	Sig.
Overall encoding	Posttest	7.155	2, 92	.001*
Shallow feature encoding	Posttest	5.998	2, 92	.004*
Deep feature encoding	Posttest	7.655	2, 92	.001*

* $p < .05$

Table 14

Results of Levene's test with combined error conditions

Variable	Measure	*F*	*df*	Sig.
Overall encoding	Posttest	14.406	1, 93	.000**
Shallow feature encoding	Posttest	11.960	1, 93	.001*
Deep feature encoding	Posttest	15.424	1, 93	.000**

* *p* < .05; ** *p* < .001

In response to these assumption violations, the study uses a Welch ANOVA of change scores to examine differences in encoding outcomes across conditions. I calculated the change scores by subtracting pretest performance from posttest performance on each of the encoding variables. The resulting change scores for the three-condition analysis is displayed in table 15. Change scores for the two-condition analysis with combined error conditions are displayed in table 16.

Table 15

Encoding performance change scores (3 conditions)

	Error anticipation (n = 41)		Incorrect worked examples (n = 35)		Correct worked examples (n = 19)	
Variable	Mean	SD	Mean	SD	Mean	SD
Overall encoding	.09	.14	.09	.14	.01	.26
Shallow encoding	.09	.15	.09	.15	.02	.26
Deep encoding	.10	.14	.10	.15	.00	.28

Table 16

Encoding performance change scores, combined error conditions

	Combined error conditions (n = 76)		Correct worked examples (n = 19)	
Variable	Mean	SD	Mean	SD
Overall encoding	.09	.14	.01	.26
Shallow encoding	.09	.15	.02	.26
Deep encoding	.10	.15	.00	.28

Data Analyses

As mentioned above, the data was analyzed using a Welch ANOVA on change scores for overall encoding as well as encoding performance on shallow and deep problem features. Analysis using three conditions found no significant difference on the overall encoding exercise as well as on shallow and deep problem features. Students in the error anticipation, incorrect worked examples, and correct worked examples experience similar changes in encoding performance. I accept the null hypothesis that there is no difference in the change in encoding performance between the error anticipation, incorrect worked examples, and correct worked examples conditions. The results of the Welch ANOVA on these data are displayed in table 17.

Table 17

Welch's one-way ANOVA of encoding change scores, 3 conditions

Variable	F	p
Overall	.869	.427
Shallow problem features	.630	.538
Deep problem features	1.080	.349

* $p < .05$

To further explore possible condition-based effects, I again examined whether there was a difference in encoding performance between students who were exposed to errors and students who received traditional instruction with correct worked examples. This analysis also found no significant difference between the combined error and correct worked examples condition. Students who worked with errors produced similar changes in encoding ability as those who worked with correct worked examples exclusively. I accepted the null hypothesis that there was no difference in change in encoding performance between the two conditions. Welch ANOVA results are given in table 18.

Table 18

Welch's one-way ANOVA of encoding change scores, combined error conditions

Variable	*F*	*p*
Overall	1.749	.201
Shallow problem features	1.227	.281
Deep problem features	2.197	.153

Research Question 3

Does including errors affect the ability and type of error predicted relative to instruction with correct worked examples?

Initial Analyses

Of the 105 students enrolled in the study, 99 completed both the pretest and the posttest. Their answers to the error anticipation items were used in this analysis. To begin analyzing student responses to the error anticipation items on the pretest and posttest, I first graded each response as correct or incorrect to analyze whether there were any differences in overall performance across the conditions. There were two error anticipation items on each, so a student could only score a 0, 1, or 2 on both tests. Subsequent analyses categorized the error by type;

however, the scoring in this analysis accounted for the number of correct responses regardless of classification. For a student to receive a score of 2, they would have to anticipate two errors that are reasonable in the context of solving the equation $5x - 2 = 8$. An example of two such responses would be "subtracting 2 from both sides instead of adding" or "combining the $5x$ and the 2". A response would be considered incorrect if the student did not answer or provided an answer that was not a reasonable error, for example showing a correct second step or provided the correct solution to the equation.

Before collapsing the individual class sections into conditions, the individual section scores were compared to ensure equal scores at pre-test. Results of independent-samples t-tests found no significant difference for the error anticipation sections, t(33.746) = -1.327, p = .193; and incorrect worked examples sections, t(42) = -1.139, p = .261. I accepted the null hypothesis that there was no difference between the sections at pretest and collapsed them into conditions. In advance of ANCOVA, the data was found to violate assumptions of homogeneity of variance at posttest [Levene's $F(2, 96) = 14.91$, $p < .001$]. Once again, change scores were computed to facilitate use of Welch ANOVA. Performance on the overall error anticipation items broken out by condition, including change scores, are displayed below in table 19.

Table 19

Error anticipation performance, 3 conditions

	Error anticipation (n = 44)		Incorrect worked examples (n = 36)		Correct worked examples (n = 19)	
Variable	Mean	SD	Mean	SD	Mean	SD
Pretest	1.75	.44	1.19	.89	1.26	.87
Posttest	1.73	.54	1.11	.89	1.05	.97
Change scores	.00	.68	-.08	1.11	-.21	1.08

I coded student responses according to the error themes utilized by McCann and Booth (2014) including arithmetic, order of operations, like/unlike terms, equals sign, negative sign, variables, other reasonable, and unreasonable/incorrect/omitted response. To check my coding for reliability, I enlisted one of the participating teachers to code a sample of 52 (approximately 26% of total student responses) student responses using those categories and tested those responses against mine. Cohen's κ analysis found substantial agreement between the two sets of classifications, $\kappa = .714, p < 0.01$ (McHugh, 2012). Table 20 provides examples of student responses and their classification. Table 21 displays error incidence at pretest across the three conditions whereas table 22 displays the same data for posttest.

Table 20

Examples of student error anticipation responses for $5x - 2 = 8$ *according to theme*

Theme	Student response
Arithmetic	"that he/she might subtract 2 to the wrong number"
	"multiply it by 2"
Order of operations	"divide by 5 first"
	"starting out with the $5x$ instead of the 2"
Like/unlike terms	"they could add unlike terms"
	"subtracting the x number (5x) by 2"
Equals sign	"forgetting to divide on both sides"
	"forgetting to add 2 to both sides"
Negative sign	"-2 from both sides"
	"they subtract the -2 instead of adding"
Variables	"$5x - 2 = 3x$ $(3x = 8)$"
	"they would multiply with x"
Other reasonable	"only do one step of the problem"
	"solve the equation"
Unreasonable/incorrect	"they might just do $5x - 2 = 8$ and leave it just like that"
	"I'm not in 7^{th} grade I don't know"

Table 21

Pretest error theme incidence, 3 conditions

	Error anticipation (n = 44)		Incorrect worked examples (n = 36)		Correct worked examples (n = 19)	
Variable	Mean	SD	Mean	SD	Mean	SD
Unreasonable/incorrect	.23	.42	.80	.89	.68	.89
Arithmetic	.00	.00	.00	.00	.11	.32
Order of operations	.30	.51	.17	.45	.16	.37
Like/unlike terms	.34	.48	.14	.35	.37	.60
Equals sign	.09	.29	.17	.51	.00	.00
Negative sign	.45	.55	.28	.45	.16	.37
Variable	.50	.59	.44	.50	.47	.61
Other reasonable	.07	.25	.00	.00	.00	.00

Table 22

Posttest error theme incidence, 3 conditions

	Error anticipation (n = 44)		Incorrect worked examples (n = 36)		Correct worked examples (n = 19)	
Variable	Mean	SD	Mean	SD	Mean	SD
Unreasonable/incorrect	.30	.55	.94	.79	.74	.81
Arithmetic	.07	.25	.06	.23	.16	.37
Order of operations	.25	.49	.14	.35	.16	.37
Like/unlike terms	.36	.53	.17	.38	.21	.42
Equals sign	.09	.29	.14	.42	.05	.23
Negative sign	.48	.50	.25	.50	.16	.37
Variable	.43	.50	.36	.64	.32	.48
Other reasonable	.07	.25	.00	.00	.00	.00

To simplify analysis, I consolidated error themes to create shallow and deep error categories. Procedural errors include arithmetic, order of operations, and other reasonable errors. Deep errors encompass the error themes of like/unlike terms, equals sign, negative sign, and variables. The resulting statistics for the three conditions are displayed below in table 23.

Table 23

Posttest consolidated error theme incidence, 3 conditions

	Error anticipation (n = 44)		Incorrect worked examples (n = 36)		Correct worked examples (n = 19)	
Variable	Mean	SD	Mean	SD	Mean	SD
Pretest procedural error	.30	.51	.17	.45	.26	.45
Posttest procedural error	.32	.52	.19	.40	.32	.58
Pretest conceptual error	1.45	.59	1.03	.84	1.00	.82
Posttest conceptual error	1.43	.66	.92	.81	.74	.81

For subsequent analysis, I again created a single combined error condition across the measures used to answer this research question. I first examined whether there was a significant difference between conditions on students' overall error anticipation score. An independent-samples t-test found a significant difference at pretest between students in the error anticipation and incorrect worked examples conditions on their overall error anticipation scores, $t(48.731) = 3.426$, $p = .001$, and their conceptual error anticipation scores, $t(78) = 3.141$, $p = .002$. There was no significant difference between the conditions on procedural error anticipation, $t(77.794) = 1.173$, $p = .233$. As a result, I rejected the null hypothesis that there was no difference across all of the sections. At pretest, students in the error anticipation sections were better able to predict possible errors than those in the incorrect worked examples sections. However, subsequent ANCOVA assumption testing found that there were no violations of homogeneity of variance for conceptual error anticipation in the two condition [Levene's $F(1, 97) = .018$, $p = .894$] and three condition [Levene's $F(2, 96) = .534$, $p = .588$] analyses. There were also no such violations for procedural error anticipation in the two condition [Levene's $F(1, 97) = .447$, $p = .505$] and three condition [Levene's $F(2, 96) = 2.108$, $p = .139$] analyses. Due to these results, ANCOVA remains appropriate for testing difference in error anticipation as it accounts for differences at pretest (Howell, 2002). Table 24 and 25 display the error theme incidence for the two-condition analysis at pretest and posttest, respectively. Table 26 displays consolidated error incidence for the two-condition analysis prior to change score calculation.

Table 24

Pretest error theme incidence, combined error conditions

	Combined error conditions (n = 80)		Correct worked examples (n = 19)	
Variable	Mean	SD	Mean	SD
Unreasonable/incorrect	.49	.73	.68	.89
Arithmetic	.00	.00	.11	.32
Order of operations	.24	.48	.16	.37
Like/unlike terms	.25	.44	.37	.60
Equals sign	.13	.40	.00	.00
Negative sign	.38	.51	.16	.37
Variable	.48	.55	.47	.61
Other reasonable	.04	.19	.00	.00

Table 25

Posttest error theme incidence, combined error conditions

	Combined error conditions (n = 80)		Correct Worked Examples (n = 19)	
Variable	Mean	SD	Mean	SD
Unreasonable/incorrect	.59	.74	.68	.89
Arithmetic	.06	.24	.11	.32
Order of operations	.20	.43	.16	.37
Like/unlike terms	.28	.48	.37	.60
Equals sign	.11	.36	.00	.00
Negative sign	.38	.51	.16	.37
Variable	.40	.56	.47	.61
Other reasonable	.04	.19	.00	.00

Table 26

Consolidated error theme incidence, combined error conditions

	Combined error conditions (n = 80)(n = 80)		Correct worked examples (n = 19)	
Variable	Mean	SD	Mean	SD
Pretest procedural error	.24	.48	.26	.45
Posttest procedural error	.26	.47	.32	.58
Pretest conceptual error	1.26	.74	1.00	.82
Posttest conceptual error	1.20	.77	.74	.81

Data Analysis

As mentioned previously, analysis of overall student error anticipation performance was complicated by violations of the assumption of homogeneity of variance. To adjust for this, change scores were computed to facilitate a one-way Welch ANOVA. Results from this analysis found no significant difference in change scores across conditions for both the three-condition analysis, *Welch's F*(2, 41.82) = .33, $p = .719$, and two-condition analysis, *Welch's F*(1, 24.11) = .417, $p = .525$. Students in the error anticipation, incorrect worked examples, and correct worked examples conditions produced similar scores on the error anticipation items. As a result, I accept the null hypothesis that there is no difference in error anticipation performance across conditions for both analyses.

To examine differences in the anticipated error theme at posttest, I used ANCOVA with pretest error theme serving as the covariate. Results for the three-condition and two-condition analyses are displayed below in table 27. Using these results, I accept the null hypothesis that there is no difference in procedural error anticipation across the conditions for the three-condition and two-condition analysis. However, I reject the null hypothesis that there is no difference in conceptual error anticipation across condition for both the three-condition and two-condition analyses.

In the three-condition analysis, post hoc Bonferroni corrections revealed that students in the error anticipation sections produced more conceptual errors (M = 1.41, SD = .12) than those in both the incorrect worked examples sections (M = .93, SD = .13) (p = .021) and correct worked examples section (M = .76, SD = .17) (p = .007). However, there was no significant difference in conceptual errors anticipated between the incorrect worked examples and correct worked examples conditions (p = 1.000). In the two-condition analysis, students who were exposed to errors anticipated more deep errors that those who worked exclusively with correct worked examples.

Table 27

ANCOVA results for error anticipation performance on consolidated error types

Analysis	Variable	df	F	*p*	Partial η^2
Three-condition	Procedural errors	2, 95	.48	.619	.010
	Conceptual errors	2, 95	6.22	.003*	.116
Two-condition	Procedural errors	1, 96	.15	.699	.002
	Conceptual errors	1, 96	4.50	.036*	.045

* $p < 0.05$

Correlations

Initial analyses included an exploration of possible correlations between the measures used across the research questions for possible relationships that might be explored in future research. As these correlations are not central to the study, there are no additional analyses for any significant relationship. However, the results will be discussed in Chapter 5. Table 28 displays these correlations.

Table 28

Correlations

Variables	1	2	3	4	5	6	7	8
1. Pretest total	1	.635**	.879**	.295**	.302**	.239*	.091	.297**
2. Pretest procedural		1	.191	.310**	.298**	.280**	.076	.075
3. Pretest conceptual			1	.181	.198	.128	.063	.335**
4. Encoding total				1	.969**	.899**	.011	.102
5. Shallow encoding					1	.765**	.007	.113
6. Deep encoding						1	.008	.071
7. Procedural error anticipation							1	.026
8. Conceptual error anticipation								1

* .Significant at the 0.05 level (2-tailed)

**.Significant at the 0.01 level (2-tailed)

CHAPTER 5

DISCUSSION

In this chapter, I will summarize and discuss the findings from the study, describe implications of the findings including how the results from this study add to the literature on using errors in classroom settings, and note future directions. Recent research has demonstrated that studying correct and incorrect examples can help strengthen mathematical knowledge (Booth et al., 2013; Booth et al., 2016). The current study examined the relative impact of error anticipation exercises compared to the use of incorrect examples in Algebra I classrooms.

The study tested the hypothesis that students exposed to error anticipation exercises would experience more growth over the course of the study in the conceptual and procedural knowledge, encoding performance, and error anticipation than those students in the correct and incorrect worked examples conditions. Although the study hypotheses were not supported, the findings suggest that including errors with correct examples is more effective than using correct examples alone. Specifically, the study found that students exposed to errors tended towards greater gains in procedural knowledge than those who examined only correct worked examples. Those students also anticipated more deep errors than those who only examined correct worked examples. Moreover, the findings suggest that error anticipation might be a viable alternative for teachers when worked example materials on a topic are unavailable, although more research is needed. The results are discussed in greater detail below.

Findings

Findings Related to Conceptual and Procedural Knowledge

The study hypothesized that students in the error anticipation condition would demonstrate greater gains in conceptual and procedural knowledge than those in the incorrect or correct only conditions. Previous work has demonstrated that studying incorrect worked examples is superior to correct worked examples (Booth & Oyer et al., 2015; Booth & Cooper et al., 2015). The generation effect suggests that as students improve background knowledge, practitioners should transition away from worked examples to problems that ask students to generate responses (Slamecka & Graf, 1978; Chen et al., 2015).

However in the current study, when controlling for pretest knowledge, there were no significant differences at posttest using a three-condition analysis. One explanation is that the study employed correct worked examples across all conditions, which have been prove effective across a wide body of research (Sweller & Cooper, 1985; Lefevre & Dixon, 1986; Grobe & Renkl, 2007; Booth et al., 2013). It is also important to note that the study employed a relatively low level of exposure to error anticipation. Students may need more time anticipating errors to counteract expertise reversal (Paas et al., 2003) and demonstrate more conclusive evidence about any effects above and beyond correct worked examples.

Another possible explanation for this finding is that the students demonstrated some prior knowledge due to their exposure to equation solving in middle school. Both the Common Core and Pennsylvania State Standards for eighth-grade mathematics includes an assessment anchor for solving linear equations, and the eligible content descriptors include such elements as writing, graphing, and solving equations, interpreting the solution to an equation or system of equations, and solving a system of linear equations in two variables (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010; Pennsylvania Department of Education, 2014). As mentioned previously, correct worked example study is generally effective for learning to solve equations (Paschler et al., 2007). However, prior exposure to equation solving likely led to the average overall pretest and posttest scores of 39% and 40%, respectively. Booth and Oyer et al. (2015) have shown that students with moderate prior knowledge tend to demonstrate less of a benefit from worked example study than those with lower prior knowledge. Examining correct worked examples remained the primary instructional tool in this study, which might have prevented this particular cohort of students from demonstrating any significant findings.

It is possible that this particular cohort of students was well positioned to benefit from more time working with errors. Results from the study support this suggestion. When the error conditions were collapsed and compared to the correct worked example condition, there was a difference approaching significance on procedural items with a small-to-moderate effect size. Students who either anticipated errors or examined incorrect worked examples tended to outscore students who were only exposed to correct worked examples. This finding is consistent with previous work that exposed students to errors in practice problems (Booth et al., 2013; Booth et al., 2016), but it builds on that work by suggesting further study into error anticipation as a possible alternative to incorrect worked examples should materials not be available for a particular topic. Incorrect worked examples, despite their proven effectiveness (Booth et al., 2013), are not commonly found in curriculum materials. Creating useful incorrect worked examples for a specific topic requires time and experience, both of which might be limited for practicing teachers (Murphy, 2015). Error anticipation exercises can be generated by simply picking an existing problem and querying the class about potential errors.

Therefore, error anticipation exercises informs the literature regarding errors in two ways. For one, they provide an opportunity to maintain student attention toward errors in topics where curriculum materials lack incorrect worked examples. At the same time error anticipation builds upon previous research regarding the generation effect, which has focused on correct solution methods, by introducing the idea of generating possible error paths.

Findings Related to Encoding

Unlike prior work (McNeil & Alibali, 2004; Booth & Davenport, 2013), this study used a pre- and posttest design for the encoding exercises to determine if there was any change in correct encoding performance across conditions. The study hypothesized that students in the error anticipation condition would better encode problem features than students in the other conditions. The students in the error anticipation condition were expected to develop stronger schemas due to the transition to generation exercises, resulting in less working memory required to process and internalize problem structures (Sweller, 1988; 2011). The three-condition analysis that compared error anticipation to correct and incorrect worked examples found no difference in encoding performance. Secondary analysis collapsed the combined error conditions and failed to find a significant difference between them as well.

Working with errors did not affect students' ability to encode problem features. The lack of a significant difference might be explained by differences from previous work in data analysis. Booth and Davenport (2013) coded errors in student encoding whereas this study measured their performance in simply replicating the equation. Employing this analysis makes it difficult to compare this study to the work of Booth and Davenport (2013). Subsequent work that employs encoding should consider analyzing and classifying student encoding errors that is more in line with previous research.

As discussed above, the minimal level of intervention might also explain the lack of a significant difference. Further, administering the encoding exercise to an entire class at once presented opportunities for students to violate protocol, which is common when extending interventions into the classroom (Murphy, 2015). This could have jeopardized the validity of the results.

Findings Related to Error Anticipation

The study hypothesized that students in the error anticipation would produce more errors overall with a greater focus on deep versus shallow errors. The proposed mechanism was the expertise reversal effect (Grobe & Renkl, 2007) whereby the transition to open-ended error anticipation would foster deeper schema development that would extend student errors beyond shallow computational mistakes.

Although the study found no significant difference between the conditions on overall anticipation performance, there was a significant difference in deep error anticipation in both the two- and three-condition analyses. Students in the error anticipation condition anticipated more deep errors at posttest (1.43) than those in the incorrect worked examples (.92) and correct worked examples (.74) conditions. When collapsed to a two-condition analysis, students exposed to errors anticipated 1.20 deep errors versus .74 for those who worked exclusively with correct worked examples.

At face value, these results come as no surprise, as students who practice either anticipating errors or who examine errors via incorrect worked examples in class should reasonably be expected to outperform those who did not. However, analysis of pre- versus post-test scores show that students in all three sections experienced at least some decrease. Those in the error anticipation condition only experienced a decrease of .02, versus decreases of .11 and .26 for students in the incorrect and correct worked examples sections, respectively. None of these differences were statistically significant. Conceptual errors include like/unlike terms, equals sign, negative sign, and variables. Students in the error anticipation condition demonstrated no change in conceptual errors anticipated. These unclear results make it difficult to draw clear conclusions about any effect on student performance on error anticipation items.

One possible explanation is that anticipating errors in class helped students in the error anticipation condition maintain their focus on errors that involve deep concepts. Participating in error anticipation likely strengthened those schema employed by the process, possibly facilitating some level of automaticity in the process (Sweller, 2011). To provide a clearer picture of error anticipation performance, future studies could consider data beyond the pre- and posttest item via student work samples.

Findings Related to Correlation

The findings discussed in this section do not directly address any research question and are presented to explore the data for potential avenues of inquiry (Slavin, 2007). A variety of significant correlations were found at pretest. The correlations demonstrated that higher pretest scores on conceptual items were associated with greater deep error anticipation. The relationship between conceptual knowledge and deep errors makes sense, as students need strong conceptual knowledge to be able to assess what constitutes a conceptual error.

The analysis also found that higher pretest scores on procedural items were correlated with higher encoding scores overall and with both shallow and deep features. This suggests that learners with stronger prior procedural knowledge are more adept at internalizing problem structures. These results appear to conflict with those of Booth and Davenport (2013), who found a relationship instead between conceptual knowledge and encoding ability. However, differences in the study instruments might account for this contrast. Their study assessed procedural knowledge with one- and two-step equations whereas the instrument for this study included expressions and equations with distributive property requiring three or more steps. Further, the instrument in Booth and Davenport (2013) was concentrated on the concepts of equivalence and like/unlike terms. The instrument for the current study included similar conceptual items but also extended beyond the scope of equation solving, for instance asking students to identify what each piece of an equation could represent within the context of a problem as well as the type of function associated with a graph, table, equation, etc. These differences and the resultant score variations could have contributed to the correlations produced by this study. Replication could provide insight into the nature of any relationship between procedural knowledge and encoding.

Limitations and Implications for Future Research

In this section I will discuss limitations of the study and suggest potential implications for future research based on the findings. This section will focus on the overarching implications that have been made apparent across the research questions, as well as presenting additional insights that have not been raised previously in the discussion.

The most obvious limitation of this study is in the ability to generalize results across all students in Algebra 1. The study used a sample of convenience at a relatively high-performing suburban high school. Prior to the study, subjects likely have stronger schema than a similar sample drawn from a school in an area with lower-socioeconomic status. As a result, all students may not be prepared to engage in generation activities without encountering extraneous cognitive load. Further, this study works exclusively with first-time Algebra students in ninth grade. As previously described, these students represent the majority of Algebra students who require the most support. A cohort of more-advanced students might find the activities to be redundant and experience expertise reversal.

Future research should consider testing the effectiveness of error anticipation exercises in a wider variety of classrooms. This could include drawing samples from Algebra 1 classrooms in a lower-socioeconomic status school as well as enlisting 8th grade sections, which are generally treated as accelerated cohorts. Doing so will allow examiners to better generalize results across all learners.

Statistical analysis in this study was, at times, complicated by the subjects' previous exposure to units on problem solving. Subsequent studies should explore topics further into Algebra I. The Pennsylvania state standards for Algebra I include topics such as functions, linear inequalities, systems of equations and inequalities, properties of exponents, radicals, polynomials, rational functions, probability, statistics, and data displays (Pennsylvania Department of Education, 2014). It is reasonable to assume that students in a ninth-grade Algebra I class have had little to no exposure to these topics. Examining a content area where participants have had minimal exposure would provide a clearer baseline from which to examine growth across conditions.

Finally, selecting a more advanced topic would move the study further into the school year. This study was conducted within the first four weeks of the school year on ninth-grade Algebra I classes. This time of year finds teachers attempting to establish norms (Goff, Kam, & Kraszewski, 2015), which could potentially impose extraneous demands on learners (Sweller, 1994). Further into the school year, it is reasonable to assume that the classroom environment is more orderly and instruction much more efficient. As such, future research should consider the timing of the study.

A second limitation, as discussed previously, could be in the interpretation of encoding results due to possible participant breaches of protocol. The encoding exercise was administered to full classrooms. To ensure fidelity, the students were instructed to leave their pencils on their desks until the equation disappeared from the screen. However, there were still instances where students had to be reminded to wait to pick up their pencil. This may have affected the validity of the results, which found no significant differences in both a two- and three-condition analysis. Future research that attempts to examine encoding across a relatively large sample should consider adjustments to the procedures to ensure participant fidelity. Individual testing would be too onerous without sizeable resources. A possible adjustment would be to administer the encoding instrument to groups of approximately 10 students at a time while administering the pretest/posttest instrument in another room. Upon completion of the encoding exercise, those students would join their cohorts and take the pretest/posttest. As the pretest/posttest takers finish they can be directed to the encoding exercise room that would restart when an adequate number of students are present. This requires some coordination with teachers in nearby rooms but it is feasible, as any class period during the school day should find several teachers on prep with available classrooms.

A third limitation that stretches across multiple research questions in the study is in the level of dosage. As previously mentioned there were a limited number of interventions across the instructional unit, approximately once every two days across ten instructional days. Future research could consider increasing dosage of error anticipation and/or incorrect worked examples. The majority of measures found no significant difference between error anticipation and incorrect worked examples conditions, though there were results that were either significant or approaching significance. This suggests that exposure to errors might provide a benefit that should be investigated further. Prior work has shown that the effectiveness of research-based interventions might not always be apparent due to the complexities of transitioning to the real classroom. Classroom research by Star and colleagues demonstrated that the effectiveness of some interventions might become clearer with increased dosage (Star & Pollack, et al., 2015). If error anticipation does, in fact, increase germane cognitive load through the generation effect, then increased dosage could help researchers find a difference should it exist.

The lack of control for teacher effect in this study constitutes another limitation. The conditions were not split evenly between the teachers. One handled both error anticipation sections, and the other was assigned the correct and incorrect examples sections. As previously mentioned, the intent of the original design was to have a more balanced mixture amongst at least 3 teachers. Future research should include a mixture of condition assignments across multiple teachers to best mitigate any teacher effect (Slavin, 2007).

The design adaptations brought about by staffing changes further complicated data analysis as a result of unequal sample sizes. To hedge against the possibility of type I error, I undertook a conservative approach that employed analysis that was robust against the resulting heterogeneous variances across the conditions (Howell, 2002; Rusticus & Lovato, 2014). The use of change scores as opposed to traditional ANCOVA or repeated-measure ANOVA has been debated in the literature (Thomas & Zumbo, 2011), and the steps required to minimize type I error likely increase the possibility for type II error (Slavin, 2007). The unequal sample sizes resulted from staffing circumstances outside the control of the study, and the decision was to keep the error sections at twice the size of the correct worked examples sections to increase statistical power. Future research with equal sample sizes should provide a clearer picture of any potential differences between the conditions.

Another limitation of this study is that it did not examine how error anticipation exercises or incorrect worked examples affected learners of varying level of prior knowledge. This study was designed to provide meaningful input to practitioners who may not have the time or resources to differentiate extensively for learners of varying levels. However, such analyses are common in the research, particularly as it pertains to cognitive load theory and expertise reversal (Grobe & Renkl, 2007; Booth and Cooper et al., 2015; Booth and Oyer et al., 2015). Future research could explore how error anticipation exercises affect students of varying ability levels. For example, it might be that incorrect examples are more effective for students with lower prior knowledge, whereas error anticipation benefits students with higher prior knowledge.

The data used in generating correlations included nested data. For instance, there is the potential for teacher effect due to the design that assigned each condition to one teacher. Although multiple safeguards were put in place to minimize teacher effect, it was not accounted for in the analysis. Multilevel modeling would be an appropriate tool for analyzing such grouped data (Howell, 2002). However, the small sample size and number of groups in the study would increase uncertainty so as to make such analysis impractical (Mass & Hox, 2005). Future research should consider accounting for nested data in correlations.

Finally, this study did not include the use of open-ended generation activities without errors. The design included correct worked examples as a control group, but the generation effect has almost exclusively been concerned with generating correct responses (Slamecka & Graf, 1978; Wireberg, Lithner, Jonsson, Lilijekvist, Norqvist, & Nyberg, 2015; Jonsson, Kulaksiz, & Lithner, 2016). Although including a fourth group might be cumbersome, it would be useful to explore any difference in outcomes for students asked to generate correct responses. Specific to this study, after correct worked example instruction the teacher could put an equation on the board and use guided practice to solicit the class for the steps in solving. To mirror error anticipation exercises, which provide the opportunity to explore multiple errors, the teacher could examine different solution paths generated by the students.

Conclusions

Although this study did not find an advantage to error anticipation activities over incorrect worked examples, the results suggest that further exploration is warranted. Executing the study across classrooms in real time presented challenges in both implementation and data analysis (Murphy, 2015; Star & Pollack, et al., 2015). Notwithstanding these complications, the study produced a wide collection of measures that allowed for novel explorations and exposed some relationships, in particular those between conceptual and procedural knowledge and encoding ability, that should be explored further. Although results were not in line with the initial hypotheses, they suggest the utility of error anticipation exercises for highlighting errors in mathematics classrooms and present new lines of inquiry. It is critical that future research continue to make its way into the classroom to ensure credibility and applicability for practitioners while at the same time providing meaningful feedback to researchers regarding the appropriate implementation of measures outside of the laboratory.

REFERENCES

Alibali, M.W., Knuth, E.J., Hattikudur, S., McNeil, N.M., & Stephens, A.C. (2007). A longitudinal look at middle-school students' understanding of the equals sign and equivalent equations. *Mathematical Thinking and Learning, 9*, 221-247.

Allwood, C.M. (1984). Error detection processes in statistical problem solving. *Cognitive Science*, *8*, 413-437.

Anderson, J.R., Farrell, R., & Sauers, R. (1984). Learning to program in LISP. *Cognitive Science*, *8*, 87-129.

Ayers, P. (2006). Impact of reducing cognitive load on learning in a mathematical domain. *Applied Cognitive Psychology, 20*, 287-298.

Baroody, A., & Ginsburg, H. (1993). The effects of instructions on children's understanding of the equals sign. *Elementary School Journal*, *84*, 199-212.

Beckmann, J.F. (2010). Taming a beast of burden – on some issues with the conceptualisation and operationalisation of cognitive load. *Learning and Instruction, 20*, 250-264.

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

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

Booth, J.L., Begolli, K.N., & McCann, N.F. (2016). The effect of worked examples on student learning and error anticipation in algebra. Paper presented at the Psychology of Mathematics Education - North America Meeting, Tuscon, AZ, November 2016.

Booth, J.L., Cooper, L.A., Donovan, M.S., Huyghe, A., Koedinger, K.R., & Pare-Blagoev, E.J. (2015). Design-based research within the constraints of practice: AlgebraByExample. *Journal of Education for Students Placed at Risk, 20*, 79-100.

Booth, J.L., & Davenport, J.L. (2013). The role of problem representation and feature knowledge in algebraic equation-solving. *The Journal of Mathematical Behavior, 32*, 415-423.

Booth, J.L., Lange, K.L., Koedinger, K.R., & Newton, K.J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Leaning and Instruction, 25*, 24-34.

Booth, J.L., Koedinger, K.R., & Siegler, R.S. (2007). 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., & Koedinger, K.R. (2008). Key misconceptions in algebraic problems solving. In B.C. Love, K. McRae, & V.M. Sloutsky (Eds.), *Proceedings of the 30th Annual Cognitive Science Society* (p. 571-576). Austin, TX: Cognitive Science Society.

Booth, J.L., Oyer, M.H., Pare-Blagoev, J., Elliott, A.J., Barbieri, C., Augustine, A., & Koedinger, K. (2015). Learning algebra by example in real world classrooms. *Journal of Research on Educational Effectiveness, 8*, 530-551.

Borasi, R. (1994). Capitalizing on errors as “springboards for inquiry”: a teaching experiment. *Journal for Research in Mathematics Education, 25*, 166-208.

Bray, W.S. (2011). A collective case study of the influence of teachers’ beliefs and knowledge on error-handling practices during class discussion of mathematics. *Journal for Research in Mathematics Education, 42*, 2-38.

Catrambone, R., & Holyoak, K.J. (1990). Learning subgoals and methods for solving probability problems. *Memory & Cognition, 18,* 593-603.

Catrambone, R., & Yuasa, M. (2006). Acquisition of procedures: The effects of example elaborations and active learning exercises. *Learning and Instruction*, *16*, 139-153.

Chen, O., Kalyuga, S., & Sweller, J. (2015). The worked example effect, the generation effect, and element interactivity. *Journal of Educational Psychology*, *3*, 689-704.

Chi, M.T., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7-75). Hillsdale, NJ: Erlbaum.

Chi, M.T., Bassok, M., Lewis, M.W., Reimann, P., & Glaser, R. (1989). Self-explanations: how students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145-182.

Chi, M.T., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439-477.

Evan, A., Gray, T., & Olchefske, J. (2006). *The gateway to student success in mathematics and science*. Washington, DC: American Institutes for Research.

Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing instructional examples to reduce intrinsic cognitive load: molar versus modular presentation of solution procedures. *Instructional Science, 32*, 33-58.

Goff, P. T., Kam, J., & Kraszewski, J. (2015). Timing is everything: Temporal variation and measures of school quality (WCER Working Paper No. 2015-4). Retrieved from University of Wisconsin–Madison, Wisconsin Center for Education Research website: http://www.wcer.wisc.edu/publications/workingPapers/papers.php

Grobe, C.S., & Renkl, A. (2007). Finding and fixing errors in worked examples: can this foster learning outcomes?. *Learning and Instruction, 17*, 612-634.

Haskell, R.E. (2001). *Transfer of learning: Cognition, instruction, and reasoning*. San Diego: Academic Press.

Hiebert, J., & Wearne, D. (1986). Procedures over concepts: The acquisition of decimal number knowledge. In J. Hiebert (Ed.*), Conceptual and procedural knowledge: The case of mathematics* (p. 199-224). Hilldale, NJ: Elbaum.

Hiebert, J.C.,& Grouws, D.A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). New York, NY: Information Age.

Hmelo-Silver, C.E., Duncan, R.G., Chinn, C.A. (2007). Scaffolding and achievement in problem-based and inquiry learning: a response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist 42*, 99-107.

Howell, D.C. (2002). *Statistical Methods for Psychology, Fifth Edition*. Belmont, CA: Duxbury Press.

Jonsson, B., Kulaksiz, Y.C., & Lithner, J. (2016). Creative and algorithmic mathematical reasoning: effects of transfer-appropriate processing and effortful struggle. *International Journal of Mathematical Education in Science and Technology, 47*, 1206-1225.

Kalyuga, S., Ayers, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38,* 23-31.

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics, 12*, 317-326.

Kirschner, P.A. (2002). Cognitive load theory: implications of cognitive load theory on the design of learning. *Learning and Instruction, 12*, 1-10.

Kirschner, P.A., Sweller, J., & Clark, R.E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75-86.

Koedinger, K.R., Booth, J.L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science, 342*, 935-937.

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

Lappan, G., Fey, J. T., Fitzgerald, W. M., Friel, S., & Phillips, E. D. (2006). *Connected Mathematics 2: Say it with symbols.* Boston: Pearson Prentice Hall.

Larson, R., & Boswell, L. (2015). *Big ideas math: Algebra 1*. Boston: Houghton Mifflin Harcourt.

Leatham, K.R. (2006). Viewing mathematics teachers' beliefs as sensible systems. *Journal of Mathematics Teacher Education, 9*, 91-102.

Lefevre, J., & Dixon, P. (1986). Do written instructions need examples?. *Cognition and Instruction, 3*, 1-30.

Mass, C.J., & Hox, J.J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*, 86-92.

McCann, N.F., & Booth, J.L. (2014). What Could Go Wrong? Error Anticipation Relates to Conceptual and Procedural Knowledge in Algebra Students. Paper presented at the annual meeting of the America Educational Research Association, Philadelphia, PA.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica, 22*, 276–282.

McNeil, N.M., & Alibali, M.W. (2004). You'll see what you mean: students encode equations based on their knowledge of arithmetic. *Cognitive Science*, *28*, 451-466.

Miller, G.A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Moses, R., & Cobb, C. (2001). *Radical equations: Math literacy and civil rights.* Boston, MA: Beacon Press.

Murphy, P.K. (2015). Marking the way: school-based interventions that "work". *Contemporary Educational Psychology, 40,* 1-4.

National Academy of Sciences. (2007). *Rising above the gathering storm: Energizing and employing america for a brighter economic future*. Washington, DC: National Academy of Sciences.

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). Common Core State Standards for mathematics: Eighth grade.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the national mathematics advisory panel*, U.S. Department of Education: Washington, DC, 2008.

National Science Board. (2003). *The science and engineering workforce: Realizing America's potential.* Arlington, VA: National Science Foundation (NSB 03-69).

National Science Board. (2012). Science and engineering indicators 2012. Two volumes. Arlington, VA: National Science Foundation (Vol. 1, NSB 12-01; Vol. 2, NSB 12-02).

National Science Foundation (2007). *Asia's rising science and technology strength: Comparative indicators for Asia, the European Union, and the United States.* NSF 07-319. Arlington, VA: Author.

Oakes, J.M., & Feldman, H.A. (2001). Statistical power for nonequivalent pretest-postest designs: the impact of change-score versus ANCOVA models. *Evaluation Review, 25*, 3-28.

OECD (2010), PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science (Volume I)

Ohlsson, S. (1996a). Learning from error and the design of task environments. *International Journal of Educational Research, 25*, 419-448.

Ohlsson, S. (1996b). Learning from performance errors. *Psychological Review, 103*, 241-262.

Paas, F.G., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educational Psychologist, 38*, 1-4.

Paas, F.G, & van Gog, T. (2006). Optimising worked example instruction: different ways to increase germane cognitive load. *Learning and Instruction, 16*, 87-91.

Paas, F.G., & Van Merrienboer, J.J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: a cognitive-load approach. *Journal of Educational Psychology, 86*, 122-133.

Pashler, H., Bain, P.M., Bottge, B.A., Graesser, A., Koedinger, K., McDaniel, M., Metcalfe, J. (2007). *Organizing Instruction and Study to Improve Student Learning* (NCER 2007-2004). Washington, D.C.: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Peled, I., & Zaslavsky, O. (2008). Beyond local connections: meta-knowledge about procedures. *For the Learning of Mathematics*, *28*, 28-35.

Pennsylvania Department of Education (2014). Mathematics assessment anchors and eligible content aligned to the Pennsylvania core standards: grade 8. Harrisburg, PA.: Author.

Pesek, D.D., & Kirshner, D. (2000). Interference of instrumental instruction in subsequent relational learning. *Journal for Research in Mathematics Education, 31*, 524-540.

Pesta, B.J., Sanders, R.E., & Murphy, M.D. (1999). A beautiful day in the neighborhood: what factors determine the generation effect for simple multiplication problems?. *Memory & Cognition, 27*, 106.

Peterson, L., & Peterson, M.J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58,* 193-198.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Instruction, 12*, 61-86.

Provasnik, S., Kastberg, D., Ferraro, D., Lemanski, N., Roey, S., and Jenkins, F. (2012). *Highlights from TIMSS 2011: Mathematics and science achievement of U.S. fourth- and eighth-grade students in an international context* (NCES 2013-009 Revised). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Radatz, H. (1979). Error analysis in mathematics education. *Journal for Research in Mathematics Education*, 10, 163-172.

Radatz, H. (1980). Students' errors in the mathematical learning process: a survey. *For the Learning of Mathematics*, *1*, 16-20.

Reifer, D.M., Chien, Y., & Reimer, J.F. (2007). Positive and negative generation effects in source monitoring. *The Quarterly Journal of Experimental Psychology*, *60*, 1389-1405.

Renkl, A. (1999). Learning mathematics from worked-out examples: analyzing and fostering self-explanations. *European Journal of Psychology of Education*, *14*, 477-488.

Renkl, A. (2005). The worked-out examples priniciple in multimedia learning. In Mayer, R.E. (Ed.), *The Cambridge Handbook of Multimedia Learning.* Cambridge: Cambridge University Press.

Renkl, A., Atkinson, R.K., & Grobe, C.S. (2004). How fading worked solution steps works – a cognitive load perspective. *Instructional Science, 32,* 59-82.

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: the effects of example variability and elicited self-explanations. *Contemporary Educational Psychology, 23*, 90-108.

Resnick, L.B., & Ford, W.W. (1981). *The psychology for mathematics instruction.* Hillsdale, NJ: Erlbaum.

Richland, L.E., Stigler, J.W., & Holyoak, K.J. (2012). Teaching the conceptual structure of mathematics. *Educational Psychologist*, *47*, 189-203.

Riefer, D.M., Chien, Y., & Reimer, J.F. (2007). Positive and negative generation effects in source monitoring. *Quarterly Journal of Experimental Psychology, 60*, 1389-1405.

Rittle-Johnson, B. (2006). Promoting transfer: effects of self-explanation and direct instruction. *Child Development*, *77*, 1-15.

Rittle-Johnson, B. & Alibali, M.W. (1999). Conceptual and procedural knowledge of mathematics: does one lead to the other?. *Journal of Educational Psychology*, *91*, 175-189.

Rittle-Johnson, B., & Kmicikewycz, A.O. (2008). When generating answers benefits arithmetic skill: the importance of prior knowledge. *Journal of Experimental Child Psychology, 101*, 75-81.

Rittle-Johnson, B., Siegler, R.S., & Alibali, M.W. (2001). Developing conceptual understanding and procedural skill in mathematics: an iterative process. *Journal of Educational Psychology, 93*, 346-362.

Rittle-Johnson, B., & Koedinger, K. (2009). Iterating between lessons on concepts and procedures can improve mathematics knowledge. *British Journal of Educational Psychology, 79*, 483-500.

Rusticus, S.A., & Lovato, C.Y. (2014). Impact of sample size and variability on the power and type I error rates of equivalence test: a simulation study. *Practical Assessment, Research, & Evaluation,* 19(11).

Salden, J.R., Aleven, V., Schwonke, R., & Renkl, A. (2008). The expertise reversal effect and worked examples in tutored problem solving. *Instructional Science, 38*, 289-307.

Santagata, R. (2004). “Are you joking or are you sleeping?” Cultural beliefs and practices in Italian and U.S. teachers’ mistake-handling strategies. *Linguistics and Education, 15*, 141-164.

Schleppenbach, M., Flevares, L.M., Sims, L.M. & Perry, M. (2007). Teachers’ responses to student mistakes in Chinese and U.S. mathematics classrooms. *The Elementary School Journal, 108*, 131-147.

Siegler, R.S. (1996). *Emerging minds: The process of change in children’s thinking*. New York: Oxford University Press.

Simon, H.A., & Gilmartin, K. (1973). A simulation of memory for chess positions. *Cognitive Psychology, 5*, 29-46.

Skemp, R.R. (1978). Relational understanding and instrumental understanding. *The Arithmetic Teacher, 26,* 9-15.

Slamecka, N.J, & Graf, P. (1978). The generation effect: delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 592-604.

Slavin, R.E. (2007). *Educational Research in an Age of Accountability*. Boston, MA: Pearson.

Star, J.R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education, 36*, 404-411.

Star, J. R., Caronongan, P., Foegen, A., Furgeson, J., Keating, B., Larson, M. R., Lyskawa, J., McCallum, W. G., Porath, J., & Zbiek, R. M. (2015). *Teaching strategies for improving algebra knowledge in middle and high school students* (NCEE 2014-4333). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved from the NCEE website: http://whatworks.ed.gov

Star, J.R., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K., Gogolen, C. (2015). Learning from comparison in algebra. *Contemporary Educational Psychology, 40*, 41-54.

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cognitive Science, 12*, 257-285.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295-312.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*, 123-138.

Sweller, J. (2011). Cognitive load theory. *Psychology of Learning and Motivation, 55*, 37-76.

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.

Sweller, J, van Merrienboer, J.J., & Paas, F.G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296.

Thomas, D.R., & Zumbo, B.D. (2011). Difference scores from the point of view of reliability and repeated-measures ANOVA: in defense of difference scores for data analysis. *Educational and Psychological Measurement, 72*, 37-43.

Thorndike, E.L. (1922). *The psychology of arithmetic*. New York: Macmillan.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology, 47*, 513-539.

VanLehn, K., Jones, R.M., & Chi, M.T. (1992). A model of the self-explanation effect. *The Journal of Learning Sciences, 2*, 1-59.

VanLehn, K., & Jones, R.M. (1993). What mediates the self explanation effect? Knowledge gaps, schemas, or analogies. In M. Polson (Ed.) *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (p. 1034-1039).

Vlassis, J. (2004). Making sense of the minus sign or becoming flexible in 'negativity'. *Learning and Instruction, 14*, 469-484.

Walston, J., & McCarroll, J.C. (2010). Eighth-grade algebra: findings from the eighth-grade round of the early childhood longitudinal study, kindergarten class of 1998–99 (NCES-2010–016). Washington, DC: U.S. Department of Education.

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction, 7*, 1-39.

Wirebring, L.K., Lithner, J., Jonsson, B., Liljekvist, Y., Norqvist, M., & Nyberg, L. (2015). Learning mathematics without a suggested solution method: durable effects on performance and brain activity. *Trends in Neuroscience and Educations, 4*, 6-14.

U.S. Department of Education. (1997). *Mathematics equals opportunity*. White paper prepared for U.S. Secretary of Education Richard W. Riley. Washington, DC: Author.

APPENDIX A

SAMPLE LESSON PROTOCOL, TREATMENT GROUP

Big Ideas Math (Larson & Boswell, 2012)

Section 1-2, Solving Multi-Step Equations

Check and collect prior night's homework (5 minutes).

Warmup (5 minutes):

- Display the following problems of the electronic whiteboard and ask students to quickly solve.

Simplify the expression:

1. $(2x^2 - 6x) - (-2x^2 + 3x)$
2. $(5a^2 - a) - (2a^2 - 5a)$
3. $(-2d^2 - d) - (5d^2 - 5d)$
4. $(2h^2 + 5z) + (2h^2 + 9z)$

Allow 2-3 minutes for students to get out notebooks and begin working. Ask students to come to the board to present answers (2 minutes).

Direct Instruction (24 minutes)

- Note-taking (3 minutes): Display textbook directions for solving multi-step equations.

- Present example 1 (3 minutes):

Solve $2.5x - 13 = 2$

Solicit student responses for appropriate steps in solving equation. Work out steps for display on electronic whiteboard.

- Present example 2 (3 minutes):

Solve $-12 = 9x - 6x + 15$

Engage students in working to solve the equation, first combining like terms and then utilizing inverse operations.

- Independent practice/check for understanding (5 minutes):

Solve the following equations:

1. $-2n + 3 = 9$

2. $-2x - 10x + 12 = 18$

Cycle room and check for understanding. Select students to present work on the board.

- Present example 3 (5 minutes):

Solve $2(1 - x) + 3 = -8$

Solve equation using two different approaches. First, solve using distributive property, combining like terms, and inverse operations. Second, solve by interpreting $1 - x$ as a single entity.

Error anticipation exercise (5 minutes):

- Display following equation on the electronic whiteboard:

$$-4(2m + 5) - 3m = 35$$

Prompt students to think about and quickly write down an error another student might make in solving the equation. Allow 1-2 minutes for students to craft responses, after which class will be prompted for potential errors.

If no student responds, or no reasonable error is suggested, the teacher will provide scaffolds in the form of suggestions. These suggestions will ask the students to consider an error category. For instance, for the above problem the teacher could ask students to consider errors with like/unlike terms or negative sign errors.

When students provide a reasonable error suggestion, the teacher will ask the student to describe what that error would look like in the above problem and display it on the board. For instance, if a student suggests "negative sign error", the teacher will ask the student to describe what a negative sign error would look like in working out the problem. After displaying the error, the teacher will ask why that step is incorrect. Each time, the teacher will erase the error and solicit additional potential errors. After soliciting student errors for 2-3 minutes, the teacher will work through the problem to display the correct solution.

Independent Practice (10-12 minutes): Distribute worksheet A for section 1-2 and have students work in pairs to complete. Homework: Worksheet B for section 1-2.

APPENDIX B

SAMPLE LESSON PROTOCOL, INCORRECT WORKED EXAMPLES GROUP

Big Ideas Math (Larson & Boswell, 2012)

Section 1-2, Solving Multi-Step Equations

Check and collect prior night's homework (5 minutes).

Warmup (5 minutes):

- Display the following problems of the electronic whiteboard and ask students to quickly solve.

Simplify the expression:

1. $(2x^2 - 6x) - (-2x^2 + 3x)$
2. $(5a^2 - a) - (2a^2 - 5a)$
3. $(-2d^2 - d) - (5d^2 - 5d)$
4. $(2h^2 + 5z) + (2h^2 + 9z)$

Allow 2-3 minutes for students to get out notebooks and begin working. Ask students to come to the board to present answers (2 minutes).

Direct Instruction (24 minutes)

- Note-taking (3 minutes): Display textbook directions for solving multi-step equations.

- Present example 1 (3 minutes):

Solve $2.5x - 13 = 2$

Solicit student responses for appropriate steps in solving equation. Work out steps for display on electronic whiteboard. Make no mention of potential errors or tripping points.

- Present example 2 (3 minutes):

Solve $-12 = 9x - 6x + 15$

Engage students in working to solve the equation, first combining like terms and then utilizing inverse operations. Again, make no mention of potential errors or tripping points.

- Independent practice/check for understanding (5 minutes):

Solve the following equations:

1. $-2n + 3 = 9$
2. $-2x - 10x + 12 = 18$

Cycle room and check for understanding. Select students to present work on the board.

- Present example 3 (5 minutes):

Solve $2(1 - x) + 3 = -8$

Solve equation using two different approaches. First, solve using distributive property, combining like terms, and inverse operations. As with previous examples, make no mention of potential errors or tripping points.

Incorrect worked example (5 minutes):

- Describe and correct the error in solving the following equation:

$$-4(2m + 5) - 3m = 35$$
$$-8m - 20 - 3m = 35$$
$$-5m - 20 = 35$$
$$-5m = 55$$
$$m = -11$$

Allow time for students to examine and respond. Students should be able to identify that the negative sign was ignored when combining like terms. The teacher will ask why this step is incorrect. In the event that no student is able to pick out the error, the teacher can work through solving the initial problem on a separate board, comparing between the two worked-out examples at each step until the error is identified. The discussion will focus only on the error presented.

The teacher will then work through the correct solution on the electronic whiteboard.

Independent Practice (10-12 minutes): Distribute worksheet A for section 1-2 and have students work in pairs to complete. Homework: Worksheet B for section 1-2.

APPENDIX C

SAMPLE LESSON PROTOCOL, COMPARISON GROUP

Big Ideas Math (Larson & Boswell, 2012)

Section 1-2, Solving Multi-Step Equations

Check and collect prior night's homework (5 minutes)

Warmup (5 minutes):

- Display the following problems of the electronic whiteboard and ask students to quickly solve.

Simplify the expression:

1. $(2x^2 - 6x) - (-2x^2 + 3x)$
2. $(5a^2 - a) - (2a^2 - 5a)$
3. $(-2d^2 - d) - (5d^2 - 5d)$
4. $(2h^2 + 5z) + (2h^2 + 9z)$

Allow 2-3 minutes for students to get out notebooks and begin working. Ask students to come to the board to present answers (2 minutes).

Direct Instruction (24 minutes)

- Note-taking (3 minutes): Display textbook directions for solving multi-step equations.

- Present example 1 (3 minutes):

Solve $2.5x - 13 = 2$

Solicit student responses for appropriate steps in solving equation. Work out steps for display on electronic whiteboard. Make no mention of potential errors or tripping points.

- Present example 2 (3 minutes):

Solve $-12 = 9x - 6x + 15$

Engage students in working to solve the equation, first combining like terms and then utilizing inverse operations. Again, make no mention of potential errors or tripping points.

- Independent practice/check for understanding (5 minutes):

Solve the following equations:

1. $-2n + 3 = 9$
2. $-2x - 10x + 12 = 18$

Cycle room and check for understanding. Select students to present work on the board.

- Present example 3 (5 minutes):

Solve $2(1 - x) + 3 = -8$

Solve equation using two different approaches. First, solve using distributive property, combining like terms, and inverse operations. As with previous examples, make no mention of potential errors or tripping points.

- Present example 4 (5 minutes)

Solve the following equation: $-4(2m + 5) - 3m = 35$

Engage students in working through steps to solve. Again, make no mention of potential errors or tripping points.

Independent Practice (10-12 minutes): Distribute worksheet A for section 1-2 and have students work in pairs to complete. Homework: Worksheet B for section 1-2.

APPENDIX D

SAMPLE INTERVENTION ITEMS

Correct Worked Example

$2(1 - x) + 3 = -8$	Write the equation.
$\underline{-3} \quad \underline{-3}$	Subtract 3 from each side.
$2(1 - x) = -11$	Simplify.
$\frac{2(1 - x)}{2} = \frac{-11}{2}$	Divide each side by 2.
$1 - x = -5.5$	Simplify.
$\underline{-1} \quad \underline{-1}$	Subtract 1 from each side.
$-x = -6.5$	Simplify.
$\frac{-x}{-1} = \frac{-6.5}{-1}$	Divide each side by -1.
$x = 6.5$	Simplify.

Incorrect Worked Example

Describe and correct the error in solving the equation:

$$-2(7 - y) + 4 = -4$$

$$-14 - 2y + 4 = -4$$

$$-10 - 2y = -4$$

$$-2y = 6$$

$$y = -3$$

Error Anticipation Item

Describe an error another student might make in solving: $-2(7 - y) + 4 = -4$

APPENDIX E

PRE- AND POSTTEST INSTRUMENT

Say It With Symbols Pretest

1. Is each of the following true? Circle yes or no for each.

a.	$x+x+x+x=4x$	Yes	No
b.	$(x+1)(x+4)=x^2+5x+4$	Yes	No
c.	$7-4=4+7$	Yes	No
d.	$x^3=x+x+x$	Yes	No
e.	$(6x+3)+(5x-2)=11x-1$	Yes	No
f.	$7x=3x-10x$	Yes	No
g.	$9-4(3-6x)=-3+24x$	Yes	No
h.	$14-6x=6x-14$	Yes	No

2. Use the distributive property to rewrite each expression in expanded form:

   a. $2(x-5)$

   b. $3(4x-7)$

   c. $(x+3)(x+2)$

3. The daily concession-stand profit in dollars P depends on the number of visitors V. The manager writes the equation below to model this relationship.

$$P = 2.50V - 500$$

She uses the equation below to predict the number of visitors V based on the probability of rain R

$$V = 600 - 500R$$

Suppose the probability of rain is 30%. What profit can the concession stand expect to make?

4. Look at the following equation for profit of a fundraiser: $P = 20x - 500 - 7x$
Circle the piece of the equation that could represent each item.

		Could be represented by:					
a.	Number of people who attended	P	20	x	500	7	None of these
b.	Price of ticket to attend	P	20	x	500	7	None of these
c.	Cost to rent the room	P	20	x	500	7	None of these
d.	Number of people that can fit in the room	P	20	x	500	7	None of these
e.	Number of tickets sold	P	20	x	500	7	None of these
f.	Cost of the snacks provided for each person	P	20	x	500	7	None of these
g.	Amount of donation given by a wealthy attendee	P	20	x	500	7	None of these

5. Solve each of the equations for x, <u>without</u> making a table or graph.

a. $-2(4 - 3x) = 4$

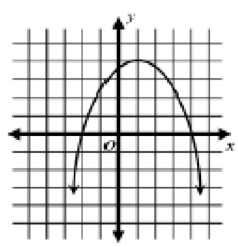
b. $x^2 + 8x + 12 = 0$

6. Think about the equation $5x - 2 = 8$

   a. What do you think is the most common mistake someone in 7th grade might make when solving the equation for x?

   b. What is another mistake you think a 7th grader might make when solving the equation for x?

7. Identify the type of function. Circle one for each:

		What type of function is it?		
a.	$y = x^2 + 6$	Linear	Quadratic	Exponential
b.	x: 2, 4, 6, 8 y: 1, 4, 7, 10	Linear	Quadratic	Exponential
c.	$y = 2^x + 6$	Linear	Quadratic	Exponential
e.	x: 3, 4, 5, 6 y: 11, 18, 27, 38	Linear	Quadratic	Exponential
f.	$y = 2x + 6$	Linear	Quadratic	Exponential
g.		Linear	Quadratic	Exponential

8. A pump is used to empty a swimming pool. The equation $w = -200(t - 6)$ Represents the gallons of water w that remain in the pool t hours after pumping starts

   a. How many gallons of water are pumped out of the pool each hour?

b. How long will it take to empty the pool?

c. How many gallons of water are in the pool at the start?

APPENDIX F

ENCODING ITEMS

1. $3x - 4 = 82$
2. $-9(2m^4 + 5) = -84$
3. $3(5p - 14) = -8p^2 + 4$
4. $89 = 2k^3 - 5(3 + 3k)$
5. $-7 + |-2m^2 + 3m| = 19$