5-2010

Misconceptions in rational numbers, probability, algebra, and geometry.

Christopher R. Rakes
University of Louisville

MISCONCEPTIONS IN RATIONAL NUMBERS, PROBABILITY, ALGEBRA, AND GEOMETRY

By

Christopher R. Rakes
B. A., University of Kentucky, 1999
M. A., University of Kentucky, 2000

A Dissertation Submitted to the Faculty of the Graduate School of the University of Louisville in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

Department of Teaching and Learning
College of Education and Human Development
University of Louisville

May, 2010

MISCONCEPTIONS IN RATIONAL NUMBERS, PROBABILITY, ALGEBRA, AND GEOMETRY

By

Christopher R. Rakes
B. A., University of Kentucky, 1999
M. A., University of Kentucky, 2000

A Dissertation Approved on

March 25, 2010

by the following Dissertation Committee:

Robert N. Ronau
Dissertation Director

William S. Bush

Maggie B. McGatha

Kathleen Moritz Rudasill

Samuel C. Stringfield

DEDICATION

To my beautiful wife and best friend Trista

ACKNOWLEDGEMENTS

No man is an island. ~John Donne

Were it not for the help, advice, friendship, and patience of so many in my life, this dissertation would certainly have never come to pass. Words cannot express the gratitude and love I have for my advisor Robert Ronau for the efforts he has made on my behalf. Dr. Ronau took the time needed to craft a program that was both attainable yet challenging, and this dissertation reflects the depth of his labors. He went beyond simply fulfilling the role of an advisor; he actively engaged me in research projects grappling with complex questions early on, and later he encouraged me to take a leadership role within several research teams. When I took up those challenges, he stood by me through every step of the journey and guided me through the various responsibilities of a research team leader. Furthermore, he modeled the very qualities he desired me to develop: love of learning, enjoyment of tackling intellectually challenging problems, a student-centered teaching style targeting the needs of every learner in every class, and resilience in the face of obstacles. He also took the time to talk me through intricate research issues to help me develop a more mature perspective of education.

I would also like to thank my committee members, William Bush, Samuel Stringfield, Kathleen Rudasill, and Maggie McGatha for their comments and suggestions throughout this project. Their advice helped shape this study into a much stronger product. Their friendship was a constant source of encouragement.

Finally, I would like to thank the many other people who have offered support, encouragement, and advice throughout my doctoral program: Thomas Tretter, Karen Karp, Todd Brown, Charles Thompson, Jeffrey Valentine, Kirsten Sundell, Jill Adelson, Murphy Shields, and Ann Larson. And, of course, without the support, encouragement, and patience of my wife Trista, this study would surely have never been completed.

“Thank you” does not even approach the love and gratitude I have for these dear friends. They have made my entire doctoral program enjoyable and rewarding.

MISCONCEPTIONS IN RATIONAL NUMBERS, PROBABILITY, ALGEBRA, AND GEOMETRY

Christopher R. Rakes
March 25, 2010

Abstract

In this study, the author examined the relationship of probability misconceptions to algebra, geometry, and rational number misconceptions and investigated the potential of probability instruction as an intervention to address misconceptions in all 4 content areas. Through a review of literature, 5 fundamental concepts were identified that, if misunderstood, create persistent difficulties across content areas: rational number meaning, additive/multiplicative structures, absolute/relative comparison, variable meaning, and spatial reasoning misconceptions. Probability instruction naturally provides concrete, authentic experiences that engage students with abstract mathematical concepts, establish relationships between mathematical topics, and connect inter-related problem solving strategies. The intervention consisted of five probability lessons about counting principles, randomness, independent and dependent event probability, and probability distributions. The unit lasted approximately two weeks.

This study used mixed methodology to analyze data from a randomly assigned sample of students from an untreated control group design with a switching replication. Document analysis was used to examine patterns in student responses to items on the mathematics knowledge test. Multiple imputation was used to account for missing data.

Structural equation modeling was used to examine the causal structure of content area misconceptions. Item response theory was used to compute item difficulty, item discrimination, and item guessing coefficients. Generalized hierarchical linear modeling was used to explore the impact of item, student, and classroom characteristics on incorrect responses due to misconceptions.

These analyses resulted in 7 key findings. (1) Content area is not the most effective way to classify mathematics misconceptions; instead, five underlying misconceptions affect all four content areas. (2) Mathematics misconception errors often appear as procedural errors. (3) A classroom environment that fosters enjoyment of mathematics and value of mathematics are associated with reduced misconception errors. (4) Higher mathematics self confidence and motivation to learn mathematics is associated with reduced misconception errors. (5) Probability misconceptions do not have a causal effect on rational numbers, algebra, or geometry misconceptions. (6) Rational number misconceptions do not have a causal effect on probability, algebra, or geometry misconceptions. (7) Probability instruction may not affect misconceptions directly, but it may help students develop skills needed to bypass misconceptions when solving difficult problems.

TABLE OF CONTENTS

TABLE OF CONTENTS

PAGE

TABLE OF CONTENTS

PAGE

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

CHAPTER 1

INTRODUCTION

Purpose Statement and Research Questions

Probability patterns often run counter to human intuition (Engel, 1970). As a result, students tend to develop misconceptions about those patterns and the mathematical concepts related to them. Some of these misconceptions fundamentally shape student understanding of mathematical patterns beyond probability (e.g., rational numbers, variables, linearity). Several researchers have proposed that probability instruction may hold the key to reducing these common misconceptions because of the abundance of concrete applications found within probability (e.g., Agnoli, 1987; Agnoli & Krantz, 1989; Bar-Hillel & Falk, 1982; Falk, 1992; Falk & Lann, 2008; Freudenthal, 1970, 1973, 1983; Shaughnessy, 1992; Shaughnessy & Bergman, 1993; Watson & Shaughnessy, 2004). Clarifying and implementing instructional tasks that are built upon the foundational nature of probability to address critical mathematical concepts in core mathematics topics offers a radical shift in how we view probability and mathematics instruction. Such a shift may create a bridge between abstract concepts and concrete applications (Freudenthal, 1983; Stone et al., 2008). The purpose of this study was to investigate the role of probability instruction as an intervention for critical misconceptions common to rational numbers, probability, algebra, and geometry by examining four research questions through a mixed methodology design:

1) Do probability misconceptions have a causal influence on algebra, geometry, and rational number misconceptions?
2) Does probability instruction reduce critical misconceptions in probability, rational numbers, algebra, or geometry?
3) Do student attitudes toward mathematics influence the emergence of errors due to misconceptions on mathematical tasks?
4) Does student metacognition influence the emergence of errors due to misconceptions on mathematical tasks?

Qualitative analysis of student responses formed the initial foundation for this study, through the analysis of error responses in order to differentiate between errors due to misunderstandings of mathematical concepts versus faulty reasoning processes. The results from this analysis were then used to code responses as indicative of misconceptions to use in the quantitative analyses.

Structural equation modeling was used to examine the causal relationship among content area misconceptions (i.e., Research Question 1). Hierarchical generalized linear modeling was used to examine the efficacy of probability instruction as an intervention for reducing misconceptions in rational numbers, algebra, geometry, and probability (i.e., Research Question 2). It was also used to analyze the impact of contextual factors on the emergence of errors due to misconceptions (i.e., Research Questions 3 and 4).

Background

Students enter high school at a time when their physical and cognitive development is at a transition point, and mathematics produces particularly strong feelings for many of these students. Students often bring preconceived notions about what it means to learn mathematics: They often have a low sense of efficacy, a great deal of anxiety, and a deep sense that much of what they learn in mathematics is irrelevant to their lives (Schumacker, Young, & Bembry, 1995). By contrast, evidence suggests that a strong command of mathematics in high school influences college success and the accessibility of many rewarding and lucrative career opportunities (National Sciences Foundation, Mathematical Sciences Education Board, 1995). Mathematics teachers in command of the nature of learning and teaching mathematics are uniquely situated to support student development (Schumacker, Young, & Bembry, 1995).

Unfortunately, evidence suggests that mathematics teaching practices have changed little to meet the needs of students in the last three decades (Hiebert, 2003). Consider the following description of traditional teaching practice:

> First, answers were given for the previous day's assignment. A brief explanation, sometimes none at all, was given of the new material, and problems were assigned for the next day. The remainder of the class was devoted to students working independently on the homework while the teacher moved about the room answering questions. The most noticeable thing about math classes was the repetition of this routine (Welch, 1978, p. 6).

The most striking feature of this description is its familiarity with current mathematics classrooms. Compare Welch's (1978) description with that of a more recent mathematics classroom:

> The typical eighth-grade mathematics lesson in the U. S. is organized around two phases: an acquisition phase and an application

> phase. In the acquisition phase, the teacher demonstrates or leads a discussion on how to solve a sample problem. The aim is to clarify the steps in the procedure so that students will be able to execute the same procedure on their own. In the application phase, students practice using the procedure by solving problems similar to the sample problem (Stigler & Hiebert, 1997, p. 18).

These two descriptions were echoed yet again by Manoucheri and Goodman (2001). Insufficient support (Tankersley, Landrum, & Cook, 2004) and minimal opportunities for professional development (Hiebert, 2003) may explain much of the inability of teachers to change their instructional practice: "Unless such opportunities are provided, teachers are asked to do the impossible – teach in new ways without having had a chance to learn them" (Hiebert, 2003, p. 18). One new way to teach that has demonstrated efficacy for helping students learn mathematics concepts is through exploratory problem solving (e.g., Mathews, 1997; Wilkins, 1993). Probability concepts inherently offer multiple opportunities for students to problem solve and explore conceptual relationships in an authentic setting (Shaughnessy & Bergman, 1993). Yet the potential of probability to meet student needs has not been realized as a result of at least two issues. First, both the intended curriculum (i.e., curriculum standards) and the enacted curriculum (i.e., what is actually taught) downplay the importance of probability relative to algebra and geometry (Mitchell, 1990; Shaughnessy, 2006; Smith, 2003). Second, teachers are less comfortable with probability due to their own lack of training and experience (Jendraszek, 2008; Shaughnessy, 1992; Swenson, 1998).

Significance of the Study

The present study responds to multiple calls for increased research about student understanding of probability concepts (e.g., Shaughnessy, 1992, 2003, 2006; Shaughnessy & Bergman, 1993; Sierpinska & Kilpatrick, 1998) and about mathematical misconceptions related to probabilistic thinking (e.g., Van Dooren, De Bock, Depaepe, Janssens, & Verschaffel, 2003) The results of this study may have direct implications for how educators view mathematical instruction.

Hypotheses

Error Responses due to Misconceptions

The mathematics knowledge instrument used in this study (Appendix M), composed of 17 released items from the National Education Assessment Program (NAEP), consisted of questions measuring algebra, geometry, rational number, and probability content. I hypothesized the types of misconceptions that might influence item responses and which distracters might indicate those misconceptions (Table 1).

Table 1
Misconception Hypotheses for each NAEP Item

Item	Correct Response	Underlying Misconception Hypothesis	Associated Responses
1	A	Absolute & Relative Comparison	C, E
2	A	Meaning of Rational Numbers: Confusion of Part-Part vs. Part-Whole	B, C, D, E
3	B	Rational Number Meaning	A
4	A	Spatial Reasoning – Interpreting arrow vs. Region	C,D
5	D	Rational Number Meaning	A, B, C
6	A	Additive vs. Multiplicative Structure	D, E
7	E	Additive vs. Multiplicative Structure	A, B, C, D
8	D	Reversal Error – Meaning of Variables	B
9	A	Spatial Reasoning: Student may choose “yes” because figure has 4 sides.	B
10	B	Spatial Reasoning: Meaning of Area – Counting Sides instead of regions.	D
11	E	Meaning of Variable – Unit Confusion, Partial Conversions	A
12	C	Additive vs. Multiplicative Structure/Coefficient Reversal	D
13	B	Rational Number Meaning	D, E
14	A	Confusion of Absolute & Relative Comparison	B
15	B	Rational Number Meaning: Part-Part vs. Part-Whole	D
16	B	Meaning of Variable	A
17	D	Absolute & Relative Comparison	C, E

Qualitative analysis of student explanations for each response was used to test these hypotheses and adjust the coding of misconception responses accordingly.

Probability Instruction as the Intervention

I conjectured that probability instruction may reduce misconceptions in rational numbers, algebra, and geometry. This hypothesis was tested using hierarchical generalized linear modeling.

Structure of Mathematical Misconceptions

Studies have indicated that rational number misconceptions and/or probability instruction hold a primary, predictive position relative to algebra and geometry misconceptions (e.g., Fuson et al., 2005; Kilpatrick et al., 2001; Lamon, 2007; Moss, 2005). A synthesis of that research, however, did not suggest which supersedes the other, nor did it demonstrate conclusively that either rational number or probability misconceptions are causal predictors of algebra and geometry misconceptions. Because probability content is inundated with rational number concepts, isolating their misconceptions is problematic without special attention to explanations of reasoning that accompany incorrect responses. Probability concepts have an advantage over rational number concepts: They naturally include concrete investigations (e.g., rolling a die, flipping a coin, examining lottery outcomes, random walks) that may help students construct meaning for abstract mathematical ideas (e.g., randomness, variation, counting principles). Based on these connections, I hypothesized that probability misconceptions act as a gatekeeper for addressing misconceptions in the other three content areas. To test this hypothesis, I compared six alternative structures (Figure 1).

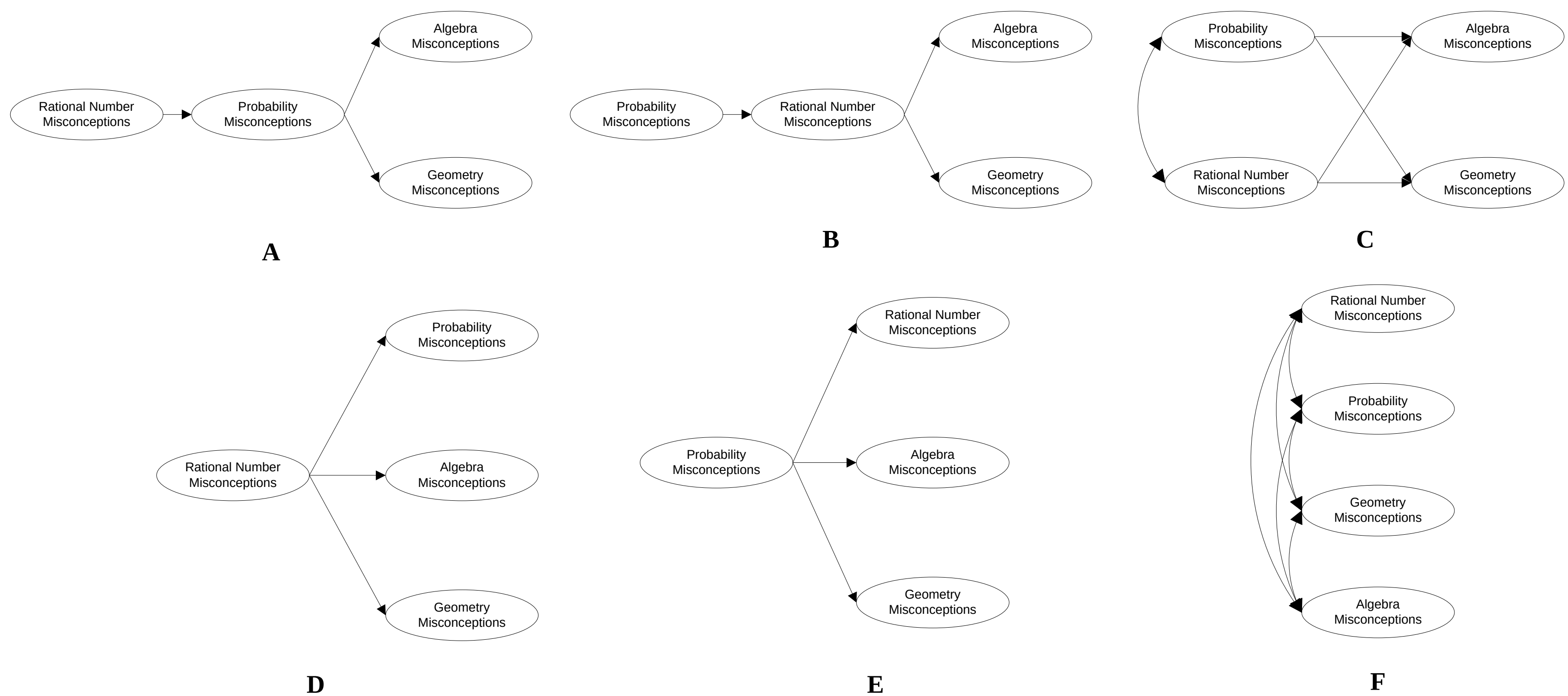


Figure 1. Six Hypothesized Structures of Mathematical Misconceptions.

Figure 1A models a relationship among content area misconceptions in which rational number misconceptions hold a primary position while probability misconceptions act as a filter on algebra and geometry misconceptions. Figure 1B reverses the relationship between rational number and probability misconceptions from Figure 1A. Figure 1C models rational number and probability misconceptions as co-varying while simultaneously exerting a causal influence on the development of algebra and geometry misconceptions. Figure 1D models the possibility that rational number misconceptions impact algebra, geometry, and probability misconceptions causally. Figure 1E reverses the role of probability and rational numbers in Figure 1D. Figure 1F models a non-causal relationship among all four content area misconceptions. I conjectured that Figure 1B or 1E would be the best fitting model.

Assumptions

Educational research is founded on beliefs about the best ways to help students learn to their fullest potential. In fact, approaches to teaching and learning cannot be separated from the underlying philosophical assumptions (Stein, Connell, & Gardner, 2008). These assumptions directly and indirectly influence the quality of learning that can take place. The three major categories of philosophical assumptions addressed in this study are epistemology, axiology, and ontology as described by Creswell, (2005) and Patton (2002).

Epistemology

Epistemology describes relationships between teachers and students, teachers and content, or students and content. Traditional views of these relationships in mathematics education consider the teacher to be an authoritative conveyer of knowledge while the students are blank slates to be filled. With very little personal interaction, traditional mathematics teaching follows a rote pattern of providing answers to homework; a brief, if any, explanation of new materials; and then students work on the assignment quietly at their desks while the teacher roams the room to answer questions (Fey, 1979), harkening back to the philosophies of Locke (Adamson, 1922) and Rousseau (1979). This traditional view considers the student and content to be completely separate, non-interacting entities. The results of this view of mathematics teaching has produced students who can inconsistently carry out mathematical procedures, have a superficial understanding of the concepts at the heart of mathematical procedures, and are unable to conduct mathematical problem solving in unfamiliar contexts (Hiebert, 2003).

In contrast, numerous researchers have suggested that students and content must interact if learning is to occur, leading to student-centered instructional approaches (e.g., Freudenthal, 1973; Hiebert & Carpenter, 1992; Hiebert & Grouws, 2007; Von Glasersfeld, 1987). Studies have found that the student-centered approach has more benefits to student learning than the traditional approach (e.g., Ford, 1977; Gregg, 1995; Hoffman & Caniglia, 2009; McMahon, 1979). Mastery learning (as in Coppen, 1976; Haver, 1978; Tenenbaum, 1986) is one example of student-centered learning: Students are tested and tutored on each topic until they achieve successful scores before proceeding to the next unit of instruction. Cooperative learning (as in Freeman, 1997; Slavin & Karweit, 1982) is another example of student-centered learning: students work in groups to facilitate peer tutoring and problem solving.

Problem solving strategies (as in Mathews, 1997; Wilkins, 1993) also provide students the opportunity to struggle with non-routine mathematical situations (as recommended by Hiebert & Grouws, 2007). Watson and Shaughnessy (2004) posited that the purposeful use of probability problem solving explorations may benefit students by providing fascinating, unique situations. The present study will not investigate the differences between teacher-centered and student-centered approaches; rather, both treatment and control groups will engage in student-centered, exploratory problem solving activities, and teacher effects will be minimized by having teachers in the study teach both a treatment and control group.

Axiology

For the present study, axiology will refer to the role of values in learning. The ability for students to learn a subject in a particular classroom from a particular teacher is greatly influenced by the alignment between student and teacher values and preferences, and learning styles (Gardner, 1987, 1989; Gardner & Hatch, 1989; Goldman & Gardner, 1989; Hatch & Gardner, 1986; Silver, Strong, & Perini, 1997). Furthermore, the value-laden nature of education constrains educators to consider the overt and covert messages being conveyed to students. Gardner (2009) identified a framework of five mental states for considering the impact on students of the transmission of values: the disciplined mind, the synthesizing mind, the creating mind, the respectful mind, and the ethical mind. These states of mind are not hierarchical; all are important. And although they may interact, these mental states do not necessarily have a causal relationship. The *disciplined mind* refers to multi-disciplinary and interdisciplinary modes of understanding, and the ability to put that intelligence into action. The *synthesizing mind* identifies the ability to pull from multiple sources and types of sources of information and combine them into a new, integrated whole. This mental frame is especially important because of the explosion of information available and the pace at which information is expanding. The *creating mind* looks beyond information and processes and innovates new processes – in the U.S., Gardner stated that the primary role of schools in relation to the creative mind is one of protection rather than cultivation. The *respectful mind* learns to value the differences in others. Although much of the cultivation of the respectful mind takes place at home, Gardner maintained that for many children, schools present the only model for respectful thinking. Therefore, he submitted that teachers must take this modeling role into account with every behavior. Children develop the *ethical mind* as they engage with questions of the type of person they want to be in the world and their place in relation to the rest of the world. These ethical thoughts, Gardner claimed, require abstract thinking that does not fully develop until adolescence. Schools play an important role in the development of this ethical frame of mind:

> Within schools, students do not literally have an occupation or a citizen's card. But for most young people, schools are the first substantial institution in which they are involved. And so it is a permissible extension to think of the vocational role of the young person as student and the citizenship role of the young person as a member of the school community. The habits of mind developed as student worker and student citizen may well help determine the ethical (or nonethical) stand of the future adult (Gardner, 2009, p. 19).

In mathematics, student values are often ignored. Students tend to value practical applicability and authentic experiences in mathematics, and the widespread absence of those qualities has resulted in motivational issues:

> Research has shown that disengagement or lack of interest is a factor in low student achievement (NCTM, 2000). Students may disengage from math because of difficulty with the subject, lack of support, or simply boredom. Students may disengage while still attending class. Many of these students believe that the math that they learn in school is not relevant to life after high school (Stone, Alfeld, & Pearson, 2008, p. 769).

The value students place on relevance is often overlooked in mathematics education in three ways: (1) Mathematics instruction often trades reasoning for rules and procedures, having the effect of separating problem solving from meaning making; (2) It emphasizes procedural understanding over conceptual understanding; and, as a result of the first two, (3) It inhibits meta-cognitive skills from being used in mathematics (Fuson, Kalchman, and Bransford, 2005). Fuson et al. (2005) proposed that the reversing of these trends will include the development of productive disposition (i.e., considering mathematics to be sensible and useful combined with a sense of self efficacy, as in Kilpatrick, Swafford, & Findell, 2001). The present study does not test these assertions. Instead, it builds on the assumption that practical applicability appeals to student values.

Ontology

Ontology describes the nature of reality. Ontological assumptions influence the meaningfulness and interpretability of research results (Patton, 2002). The ontological assumptions of the present study will be organized by responding to four questions: (1) Are mathematical concepts part of a "singular, verifiable reality and truth" or the result of multiple socially constructed realities; (2) How do people know mathematics; (3) How should mathematics be studied; and, (4) What mathematics is worth knowing?

Are mathematical concepts part of a "singular, verifiable reality and truth" or the result of multiple socially constructed realities? The present study proceeds from the basis that a single, objective reality exists for mathematics, but understanding such a reality requires students to filter it through social constructs. As a result, the nature of mathematical reality as it is understood from person to person varies. This assumption is closely tied in with how people know mathematics.

How do people know mathematics? Kant (1786/1901) proposed the importance of intuition (described as the only way human knowledge can relate to an object) to learning mathematics. "All human cognition begins with intuitions, proceeds from thence to conceptions, and ends with ideas" (p. 516). He divided intuition into two categories: empirical and pure. He defined empirical intuition as the intuition of the senses, and any object of empirical intuition as a phenomenon. He defined *sensation*, then, as the "effect of an object upon the faculty of representation" (p. 63). Empirical intution can only exist after experience, or as *posterior* intuition (p. 63). Pure intuition, by contrast, refers to the organization of objects prior to sensation. Pure intuition is therefore independent of experience: The stripping away of properties such as "substance, force, divisibility, impenetrability, hardness, color, etc." leaves two characteristics that belong to the form of the object: *extension* and *shape* (p. 64). Developing these concepts further, he defined the two objects of pure intuition, which must exist a priori and external to experience, as space and time. Mathematical conceptions proceed from intuitions: "Mathematical cognition is cognition by means of the construction of conceptions. The construction of a conception is the presentation *a priori* of the intuition which corresponds to the conception" (p. 522). In the present study, *intuition* will be used to refer to the way mathematical objects are understood. Although these intuitions are independent of training and, to a certain degree, experience, the present study assumes that natural intuition can be modified through experiences that provide cognitive dissonance (Festinger, 1957).

How should mathematics be studied? Hiebert and Grouws (2007) described two fundamental characteristics of a mathematics classroom that focuses on conceptual knowledge and relational understanding. First, explicit attention to concepts supports the development of conceptual understanding. The effect of conceptual focus has been demonstrated across research designs, teaching styles, and classroom environments. Second, teachers allow students to struggle with important concepts. By use of the term *struggle*,

> We do not mean needless frustration or extreme levels of challenge created by nonsensical or overly difficult problems. We do not mean the feelings of despair that some students can experience when little of the material makes sense. The struggle we have in mind comes from solving problems that are within reach and grappling with key mathematical ideas that are comprehendible but not yet well formed (p. 387).

Although skill efficiency and conceptual struggle are not mutually exclusive, each mode of teaching relies on a different features within the classroom. However, Hiebert and Grouws (2007) noted that studies in their review found high skill levels in a variety of class types that focused on conceptual understanding (e.g., teacher-centered versus student-centered). Learning skills in a conceptual environment versus a procedural one seems to afford students increased ability to adapt their knowledge to new situations. In the present study, the importance of conceptual focus will not be examined. Instead, both treatment and control groups will receive conceptually-based instruction.

What mathematics is worth knowing? A curriculum is generally set in place to delineate the important topics to be studied. How the term is understood varies between groups. Teachers usually consider curriculum to refer to goals or objectives, textbooks, standards documents, printed materials, lesson plans, study sheets, or tests; administrators, on the other hand, may be more interested in the material taught by teachers or commercial programs (Sinclair & Ghory, 1979). Parents may consider curriculum to mean the types of courses offered by a school. Reys and Lappan (2007) described all of these notions of expressed curriculum as the *intended* curriculum (p. 676). Sinclair and Ghory (1979) identified three other dimensions to the meaning of curriculum: expressed, implied, and emergent.

The expressed curriculum refers to “learning objectives, learning opportunities, sequence of content, and evaluation procedures” (p. 5). The expressed curriculum carries the teacher’s interpretation of the intended curriculum into the classroom. The infusion of ever-increasing content demands and pressure from standardized testing has resulted in the “mile wide, inch deep” curriculum (Schmidt, Houang, & Cogan, 2002, p. 3). Reys and Lappan (2007) found that content emphases in mathematics vary widely between states. They suggested that future revisions of state documents should include collaboration between states with a great deal of national direction.

The implied curriculum is the expression of unspoken messages through classroom policies and procedures and school culture. The implied curriculum holds special importance for the present study by its reference to “unintended learning that results because of what is included or omitted in the content that is taught” (Sinclair & Ghory, 1979, p. 6). In many classrooms, authentic mathematical experiences are omitted from the curriculum in order to move students through content more quickly (Hiebert & Grouws, 2007). Sometimes, content is deleted or minimized due to teachers’ lack of familiarity or comfort level with the content, as is often the case with probability (Shaughnessy & Bergman, 1993).

The emergent curriculum represents a response from the teacher based on formative assessment, resulting in adjustments to the expressed curriculum as needed to fill in gaps between learners and content (Sinclair & Ghory, 1979). The emergent curriculum can be considered a tool that is especially important for teachers to reduce mathematical misconceptions.

Little consensus exists in the United States about critical issues such as how to rate the importance of specific topics within a curriculum, the role of accountability testing, or the appropriate time to introduce important concepts. The present study did not attempt to resolve these issues; instead, it proceeds from the assumption that each school and teacher addresses curriculum issues differently. As a result, the random assignment was stratified to divide the effects of these differences across both treatment conditions.

Limitations

Several teachers replaced the researcher-provided conceptually-based instructional materials with procedurally-based materials for the probability intervention unit. These teachers cited several reasons for doing so (e.g., not believing that their students could handle the provided materials and discomfort with the probability material). While teacher effects were controlled across treatment and control groups, the observed magnitude of intervention effects may be reduced from a more homogenous conceptually-based intervention. Therefore, the analysis was limited to relative comparison of effects.

The testing instrument for mathematics knowledge may have also limited the analysis of the study. The literature review identified an underlying set of foundational misconceptions that were unable to be measured discretely from the distractor responses. As a result, subsequent hypotheses about the structure of these fundamental misconceptions could not be tested with these data.

Organization of the Remaining Chapters

The following chapters provide a rationale for the investigation of probability misconceptions along with the methodology, results, and conclusions of the study. Chapter 2 examines the research foundations for the present study. A synthesis of this review allowed for the development of a conceptual framework for how students learn and misunderstand mathematical ideas.

Chapter 3 provides a rationale for the research design and methodology decisions made throughout the study. These decisions included how to recruit subjects, how to assign classes to treatment groups, determination of sample sizes needed to have adequate power, appropriate analytic techniques for each research question, and how to handle missing data to maximize power while minimizing threats to validity. This chapter also includes a description of the treatment, treatment procedures, and assessment instruments.

Chapter 4 begins with descriptive statistics of the sample and the results of each test. The chapter goes on to report the results of the qualitative analysis of student responses to each item in the mathematics knowledge test and how that analysis informed the coding of error responses. The chapter then provides the statistics from the analysis of the structure of misconceptions. The chapter ends by presenting the results of the contextual factor analysis.

Chapter 5 begins by discussing how the structural analysis underscores the inadequacy of organizing misconceptions by content areas. It continues by discussing how the qualitative and structural analyses taken together suggest the need for the development of a new instrument to specifically measure misconceptions. The chapter concludes by discussing the contextual factors that influenced the production of errors due to misconceptions for each task and the effectiveness of the probability unit intervention.

CHAPTER 2

LITERATURE REVIEW

As students develop mathematical thinking and reasoning, several key stumbling blocks prevent deep conceptual learning (e.g., the transition from whole numbers to rational numbers in elementary and middle school as in Moss, 2005). These problems often persist throughout high school, adding to the difficulties of transitioning from arithmetic to algebra (Kilpatrick, Swafford, & Findell, 2001). Throughout these transition periods, students may attempt to incorporate new information into their current knowledge base without having sufficient understanding to successfully bridge the ideas (MacGregor and Stacey, 1997). Errors resulting from these misunderstandings may indicate a common set of misconceptions that affect the learning of every mathematics content area.

Defining Misconceptions

The term *misconception* has been used in research to refer to a wide range of issues, and its use has evolved through two phases (Confrey, 1987). The first phase, from the early 1970's to the early 1980's, laid the foundation for examining misconceptions as ideas that emerge from students examining problem solving situations intuitively, making decisions that appear rational yet lead to errors (Clement, 1982; Confrey, 1987). These errors often surprise educators, are difficult to eradicate, and affect a large portion of people.

In the second phase of misconceptions research, mathematics educators focused on errors rather than misunderstandings. Slip, bugs, and repair theory concentrated on procedural errors (e.g., VanLehn, 1980; 1983) while systematic errors focused more on conceptual errors.

> Systematic errors include the systematic (and inappropriate) application of familiar fragments of arguments, algorithms and definitions without any attempt to integrate across representational systems. They are common across students, and permit accurate predictions of what answers students will give to a set of well-defined problems (Confrey & Lipton, 1985, p. 40).

A recent study in Kentucky shed light on the comparative strengths of focusing on conceptual errors instead of procedural errors (McGatha, Bush, & Rakes, 2009). The study compared student achievement resulting from observed teacher assessment behavior, including addressing procedural errors and conceptual errors. Teachers who focused on deep reasoning (7th grade: a non-testing year in Kentucky) saw the highest gains in student achievement. Teachers who focused on procedural errors (8th grade: a testing year in Kentucky) obtained the least amount of growth in student achievement. Another study examining instructional strategies in algebra found that teaching methods focused on helping students develop connections between ideas produced larger and more consistent effect sizes than interventions that targeted procedural fluency (Rakes, Valentine, & McGatha, 2010).

Misconceptions versus Reasoning Errors in Secondary Mathematics

The symptomatic features of mathematical misconceptions are often discussed in literature simultaneously with reasoning errors (e.g., Falk, 1992; Kahneman & Tversky, 1972, 1973a, 1973b; Küchemann, 1978). For the purposes of this study, the following discussion does not present an exhaustive list of reasoning error types. Three types of reasoning errors appear to be related or confused with misconceptions. Some reasoning errors result from misunderstandings about ideas and connections among ideas, in which case they may indicate an underlying misconception (e.g., Clements, 1982). Other reasoning errors emerge from misunderstandings about mathematical procedures (e.g., Walker & Singer, 2007). In such cases, the present study does not consider such reasoning errors to represent misconceptions. A third type of reasoning error may arise from a combination of conceptual and procedural misunderstandings (e.g., De Bock, Verschaffel, & Janssens, 1998, 2002). In such cases, misconceptions are often difficult to parse out from these other types of error patterns. The present analysis attempts to do so by discussing error patterns as they appear in the literature in enough detail to separate heuristic reasoning errors from conceptual reasoning errors. Understanding the source of errors carries important consequences for how teachers address misconceptions:

> There is a tendency for teachers when confronted with a statement from a student that is apparently incorrect to inform the student of the error and perhaps state the correct point of view…The view that these statements are, however, not isolated beliefs that the student holds but are reflections of a more general conceptual framework leads one to be skeptical about the effectiveness of these types of local interventions; they

> do not get at the heart of the problem. An analogous situation would involve correcting the assertion that a ship would fall off the earth if it ventured too far from shore by negating the assertion or citing evidence to the contrary rather than focusing on the apparent underlying belief in a flat rather than spherical planet (Konold, 1988, p. 18).

If an error within a task occurs because of a fundamental misconception of a mathematical idea or the relationship between ideas, then directing the student's attention to a procedure-based correction within the task may be insufficient. Such an attempt to fix the error may appear successful for a specific type of task, but when students face a new, unfamiliar situation, the misconception will often reassert itself on student reasoning (Hiebert & Grouws, 2007). The same situation can occur when the misunderstandings are a combination of meaning and procedures (e.g., Fisher, 1988; Phillippe, 1992). Instead, interventions that are effective in the long term will address the lack of understanding about the meaning of the important mathematical ideas (Hiebert & Grouws, 2007).

If an error in reasoning occurs despite conceptual understanding, then addressing the error by focusing on the procedures may be effective. Focusing on the underlying meaning and reviewing how procedures relate to that meaning will, however, reinforce understanding of the structure and relationships of the mathematical ideas (Kieran, 1989, 1992, 2007). So regardless of the source of the error, focusing on the underlying meaning and connections of mathematical ideas appears to offer the longest lasting benefits for students (Skemp, 1976/2006).

These connections may be especially important for reinforcing student struggle with the specific hurdles of learning rational numbers, algebra, and geometry; conversely, the difficulties in one area often influence student ability to handle the difficulties in another area. Rational numbers, for example, may play a fundamental role in students' ability to solve algebraic and geometric problems. Misunderstandings about the connections between variable symbols and the meaning of variation may influence student capacity to understand probability concepts such as randomness.

Several key concepts from algebra, geometry, and rational numbers also appear to influence multiple facets of probability. The relationship between these misconceptions may suggest a connection between probability and these other three content areas. Such a connection may indicate that probability instruction may offer a unique inroad into addressing misconceptions about each of these areas by developing the meaning of fundamental concepts important to each topic.

Critical Misconceptions Specific to Learning Rational Numbers

Rational Number Meaning

Rational number concepts confound student mathematical understanding more than whole numbers, in part because of the multiple representations and uses of rational numbers and the major conceptual shift that is required of students when learning rational numbers (Fuson et al., 2005; Kilpatrick et al., 2001; Lamon, 2007; Moss, 2005).

> When fractions are treated as numbers in the beginning of the journey – too early on – learners often assume that the greater the denominator the greater the amount – $\frac{1}{8} > \frac{1}{7}$ because 8 > 7. Even when they begin to understand that the denominator is a divisor, and therefore the greater the number of pieces, the smaller the amount, the relationship

> of the numerator to the denominator escapes them (Fosnot & Dolk, 2002, p. 56).

Two of the most common rational number relationships have been described as "part/part" and "part/whole" (e.g., Baturo, 1994; Behr, Harel, Post, & Lesh, 1992). Part/part relationships occur when a quantity within one unit is compared to a quantity within another unit (Lamon, 1999). For example, a male/female ratio represents a part/part relationship: neither quantity represents the total number of people in the class. Part/whole relationships, on the other hand, represent the relationship between a part and a whole. For example, the male percentage of a class represents a part/whole relationship. Negotiating between part/whole and part/part relationships and the quantities represented by each may be critical to overcoming the rational number hurdle (Fosnot & Dolk, 2002):

> One third of one strip of paper is not equivalent to one third of another, shorter strip of paper. It is this relational thinking that makes fractions so difficult for children. The parts must be equivalent, but they must also be equivalent in relation to the whole (p. 56).

Behr et al. (1992) agreed with Fosnot and Dolk's connection between difficulties with fractional meaning and equivalence, elaborating on the multiplicative structure of rational number relationships:

> Fundamentally, the question of whether two rational numbers are equivalent or which is less is a question of invariance or variation of a multiplicative relation…Two rational numbers a/b and c/d, can be compared in terms of equivalence or nonequivalence by investigating whether there is a transformation of a/b to c/d, defined as changes from a

to *c* and from *b* to *d*, under which the multiplicative relationship between *a* and *b* is or is not invariant (p. 316).

The part/whole relationship is a necessary step in understanding rational numbers, but it may not be sufficient to developing meaning. Wu (2005) described the importance of rational numbers to mathematical understanding and the insufficiency of the part/whole relationship.

> The subject of fractions (which is the term I will use for nonnegative rational numbers) is known to be a main source of mathphobia. If this is not reason enough for us to teach fractions better, let me cite another one: understanding fractions is the most critical step in the understanding of rational numbers because fractions are students' first serious excursion into abstraction. Whereas their intuition of whole numbers can be grounded on the counting of fingers, learning fractions requires first of all a mental substitute for their fingers. They need to be clearly told what a fraction is. A fraction has to be a number, and so the definition of a fraction as "parts-of-a-whole" simply doesn't cut it. Students have to be shown that fractions are the natural extension of whole numbers so that the arithmetic operations +, −, x, and ÷ on whole numbers can smoothly transition to those on fractions (p. 2).

The fundamental concept of relative versus absolute size interacts with students' ability to interpret the part-whole relationship correctly. Students in elementary grades tend to mix these two comparative techniques up; this confusion persists into the high school years (Green, 1983b; Watson & Shaughnessy, 2004). Steen (2007) described interpretation as a more fundamental problem for students than the computation of rational numbers. For example, students have difficulty approximating rational number values, such as the sum of 19/20 and 23/25. Given the choices of 1, 2, 42, or 45, most eighth grade students in the U. S. chose either 42 or 45, indicating difficulties in ascribing meaning to the relationship between each part of a rational number. He observed that computers and calculators can help with many of the computational difficulties, but these tools are unable to bridge gaps in meaning. Schield (2006) found that these difficulties extend to percentage representations of rational numbers as well. Consider an example from the Schield Statistical Literacy Survey (Schield, 2002, p. 2):

> Do you think the following statements <u>accurately describe</u> the data shown in [Figure 2]?
>
> [Question] 9. 20% of smokers are Catholic
>
> [Question] 10. Protestants (40%) are twice as likely to be smokers as are Catholics (20%).

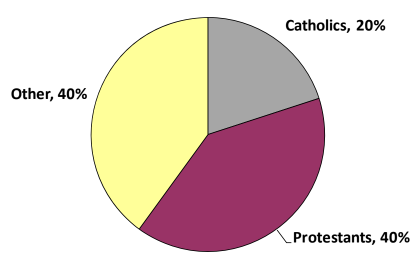


Figure 2. Part-Whole and Part-Part Comparison using Pie Chart (Schield, 2002, p. 2).

Question 9 asks students to interpret a relationship between a part (i.e., Catholics) and the given whole (i.e., Smokers). Schield (2006) reported that only 19% of the college students in his sample analyzed this relationship incorrectly. By contrast, 62% missed Question 10. This question digs deeper into student understanding of the part-whole identities. Since 20% of smokers are Catholic and 40% of smokers are Protestant, students (correctly) conclude that the number of Protestant smokers is twice that of Catholics for this sample. Question 10 reverses the logic of Question 9. Although 20% of smokers are Catholic, the graph does not indicate the converse: 20% of Catholics are not necessarily smokers, nor are 40% of Protestants (based on the chart). The difficulties students exhibited on Question 10 indicate that they have a limited understanding of the nature of part-whole relationships. Schield (2006) also found that tables reporting percentages present a similar difficulty for students, as in the following example:

> Do you think the following statements accurately describe the 20% circled [in Table 2]?
>
> Q30. 20% of runners are female smokers.
>
> Q31. 20% of females are runners who smoke.
>
> Q32. 20% of female smokers are runners.
>
> Q33. 20% of smokers are females who run (Schield, 2002, p. 6).

Table 2
Two-Way Half Table

	PERCENTAGE WHO ARE RUNNERS		
	Non-smoker	Smoker	Total
Female	50%	20%	40%
Male	25%	10%	20%
Total	37%	15%	30%

Note. From Schield, 2002, p. 6.

In Table 2, each percentage represents a part-whole ratio: the numerator (the part) represents the quantity of runners while the denominator (the whole) represents the intersection of the row and column quantities. One of the primary clues for reading this table is the lack of any 100%'s in any cell of the table: These missing values should be interpreted as meaning that each percentage represents a portion of a different quantity from the others. So, runners comprise only 50% of female non-smokers while they make up 25% of male non-smokers. The circled 20% thus represents the statement that runners include 20% of female smokers. Schield (2006) reported high error rates among college students on all four questions above: 55%, 53%, 62%, and 42% respectively. Each question examines student understanding of part/whole relationships from a different perspective. Question 30 reverses the role of the whole (i.e., female smokers) and the part (i.e., runners). Question 31 confuses smokers as part of the whole along with runners while Question 33 confuses females as part of the whole. Question 32 correctly identifies the role of each quantity.

In contrast to a two-way half table, 100% row tables and column tables must be interpreted differently.

Table 3
100% Row Table

	SEX		
	Male	Female	Total
Black	75%	25%	100%
White	50%	50%	100%
Other	40%	60%	100%
Total	50%	50%	100%

Note. From Schield, 2002, p. 6.

Table 4
100% Column Table

	College Students		
Major	Male	Female	Total
Business	60%	20%	40%
Economics	10%	50%	30%
Miscellaneous	30%	30%	30%
Total	100%	100%	100%

Note. From Schield, 2005, p. 1

In Table 3, the 100%'s in the row marginal cells indicate that each cell percentage represents a portion of the row quantity. So, females make up 25% of the Black sample while they account for 50% of the White sample. In Table 4, the column quantities now represent the whole. So, business majors account for 60% of the male sample while they account for only 20% of the female sample. Schield (2005, 2006) considered these errors to represent fundamental misunderstanding about the meaning of rational numbers and the connection between different representations of rational numbers. The table format of the questions may also have contributed to student errors, which would not necessarily represent underlying mathematical misconceptions. His analysis, however, suggested that sufficient evidence indicated mathematical misconceptions unique from difficulties with table formats.

Probability Connections to Rational Number Meaning

Rational number difficulties instill a sense of frustration and anxiety about mathematics (Gresham, Sloan, & Vinson, 1997); on the other hand, probability applications of rational numbers may provide the concrete examples students need to be able to derive meaning from these number relationships thereby reducing that anxiety.

Probability applications regularly expose students to part/whole relationships, readily providing concrete substitutes for fingers. These applications go beyond simply counting outcomes: They also ask students to examine which quantities to count for each part of the probability ratio, how to count them, and how to compare those values. Green's (1982) counter problem provides an example of how rational number relationships can be examined in a probability context (p. 20):

6 (e) Two other bags have black and white counters.

Bag J: 3 black and 1 white

Bag K: 6 black and 2 white

Which bag gives a better chance of picking a black counter?

(A) Same Chance (B) Bag J

(C) Bag K (D) Don't Know

Why?

Sixty-two percent chose C, citing the larger number of black counters as the reason for their choice. This error highlighted an underlying misconception about rational number equivalence: Students failed to recognize that the number of black marbles was not being compared, but the relationship of the number of black marbles to the whole in each bag.

In addition to interpreting the meaning of a single rational number relationship, students are often asked while studying probability to compare whether two sets of rational number relationships are equivalent, i.e., the linear proportion. Linear proportions are highly useful for solving a wide array of mathematical problems (Van Dooren et al., 2003). Linear patterns are also highly intuitive because of their simplicity (Rouche, 2003). However, Freudenthal (1983) recognized a potential misconception regarding linearity: "Linearity is such a suggestive property of relations that one readily yields to the seduction to deal with each numerical relation as though it were linear" (p. 267).

Linear Proportions

Van Dooren et al. (2003) determined that many misconceptions about numerical relationships can be traced from the overgeneralization of linearity or proportionality. They described several types of situations in which students tend to apply linear proportions although the situation actually contained an additive structure. For example, "Sue and Julie were running equally fast around a track. Sue started first. When she had run 9 laps, Julie had run 3 laps. When Julie completed 15 laps, how many laps had Sue run?" (p. 114). In their study, the subjects were elementary education teachers rather than students, yet they reported that 97% of their sample solved the problem using a proportion, $9/3 = x/15$ (computing $x = 45$), instead of $x + 6$, for a correct answer of 21. Their study revealed a striking pattern: Subjects tenaciously held to their faulty reasoning, some through four interviews focused on correcting the misunderstanding. Their study revealed some of the difficulties students have modeling quantities within a rational number relationship: Students who opted for a straight linear proportion indicated through their interviews that they believed the relationship between Julie and Sue is proportional (i.e., multiplicative), rather than additive.

Lamon (1999) offered an example of similar errors related to interpreting mathematical quantities. This example suggested that the errors are not simply overusing linearity, as Van Dooren et al. (2003) later suggested. Instead, Lamon noted that students do not only overuse proportions; they also use addition when proportions would have been appropriate. Lamon suggested that such errors stem from fundamental misapplication of meaning — students attempting to connect two mathematical relationships without truly understanding the meaning of either. Consider her example of snake growth:

> Jo has two snakes, String Bean and Slim. Right now, String Bean is 4 feet long and Slim is 5 feet long. Jo knows that two years from now, both snakes will be fully grown...At her full length, String Bean will be 7 feet long, while Slim's length when he is fully grown will be 8 feet. Over the next two years, will both snakes grow the same amount? (Lamon, 1999, p. 12).

Using an additive structure to compare the absolute growth rates, one can consider that both snakes will grow the same amount, three feet. On the other hand, comparing the relative growth rates requires considering the amount of growth in relation to the original size through a multiplicative structure. String Bean's additional three feet will be an additional 3/4 or her original length, while Slim's additional growth will only be an additional 3/5 of his original length. So, the additive comparison reveals the same amount of growth, but the multiplicative comparison reveals a different rate of growth. Each interpretation is correct within the context of answering a particular question, and both result in erroneous solutions if their meanings are confused. Such errors of meaning resulting in faulty reasoning get at the heart of mathematical misconceptions.

Probability Connections to Linear Proportions

These errors of meaning emerging from confusion about the nature of linear proportions also appear in probability. The Birthday Paradox provides a well-known example of this connection:

> If in a gathering of 50 people one asks how probable it is that there are two people with the same birthday in the room, it is nearly always the

case that this probability is grossly underestimated. The mathematician who stages this can count on a success such as only magicians can boast of, if several pairs, maybe even a triple, can be found with the same birthdays (Freudenthal, 1973, p. 587).

Several interpretations have been offered to explain students' difficulties with this issue. Kahneman and Tversky (1982) suggested that the underlying misconceptions result from the misuse of the linear proportion arising from reliance on the representativeness heuristic:

> Most students are surprised to learn that in a group of as few as 23 people, the probability that at least two of them have the same birthday (i.e., same day and month) exceeds .5. Clearly, with 23 people the expected birthdays per day is less than $\frac{1}{15}$. Thus a day with two birthdays, in the presence of 343 "empty" days, is highly non-representative, and the event in question, therefore, appears unlikely (p. 37).

The expected ratio of $\frac{1}{15}$ emerges from the expectation of equivalency: $\frac{23}{365}$ is approximately $\frac{1}{15}$, or 7%. However, counting only 23 possible matches only accounts for the potential matches of one person to the other 22. The other 22 could each have matches as well. A simple example may clarify the appropriate counting techniques: Suppose we randomly choose four people, and their birthdays are labeled A, B, C, & D. The possible matches in this scenario are A = B, A = C, A = D, B = C, B = D, and C = D. To count these matches, we see that there are three potential matches for A, two for B, and only one distinct match left for C, or 3 + 2 + 1 = 6 potential matches. If we were to add a fifth person E, the counts for A, B, and C would increase by 1, and D would now have a unique possible match with E. Thus, the general formula for counting the number of possible matches for n randomly chosen people is 1 + 2 + 3 + … + (*n* – 1), an arithmetic series with a constant increase of one unit per term. The formula for the sum S_n of an arithmetic series is given in Equation 1.

$$S_n = \frac{n(t_1 + t_n)}{2} \tag{1}$$

In this series, the "*n*th" term is actually the "(*n* – 1)th" term, and the first term has a value of one, so the formula after substitution becomes Equation 2.

$$S_{n-1} = \frac{(n-1)(1+n-1)}{2} = \frac{n(n-1)}{2} = \frac{1}{2}n^2 - \frac{1}{2}n \tag{2}$$

Another way to arrive at Equation 2 begins by using the formula for combinations to compute the number of ways to choose any two people from a group of size *n*, or ${}_nC_2$. Using substitution, we arrive at Equation 3, which simplifies to Equation 2 by canceling out the (*n* – 2)! term.

$${}_nC_2 = \frac{n!}{2!(n-2)!} = \frac{n \bullet (n-1) \bullet (n-2)!}{2 \bullet (n-2)!} \tag{3}$$

Rather than being linear, the pattern of counting potential matches follows a quadratic pattern (Figure 3).

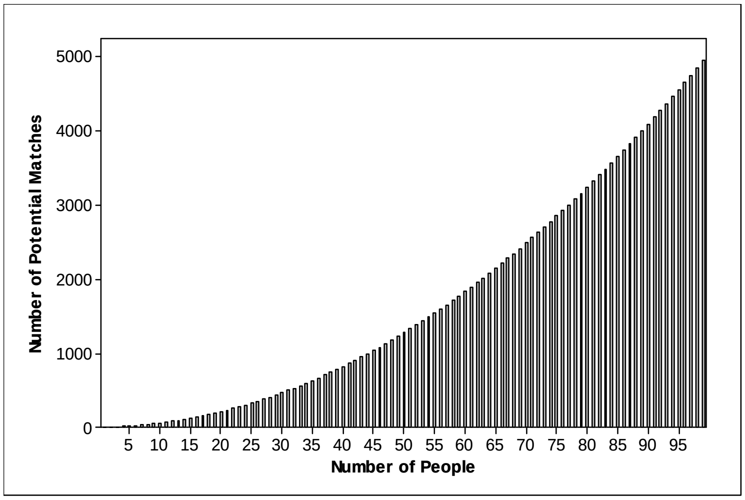


Figure 3. Number of Potential Birthday Matches for Groups of 2 – 100 people.

Returning to the original examples discussed by Freudenthal (1973) and Kahneman and Tversky (1982), we can readily compute that 50 people have 50*49/2 = 1225 possible matches, and 23 people have 23*22/2 = 253 possible matches. These numbers cannot be used as either the numerator or denominator of the desired probability: 1225 is larger than 365, and probabilities larger than one are impossible. Using 253/365 is tempting, however, to do so assumes once again that the change in probability is linear, which eventually would lead to probabilities larger than one. To examine this problem closer, we turn to the Fundamental counting principle. Since only one day out of each year can provide a successful match for any randomly chosen person, we can conclude that the probability of no matches is 364 days out of 365 and that each potential match is independent of the others. The Fundamental counting principle stipulates that the probabilities should be multiplied together. Equation 4 demonstrates this calculation for four people (six potential matches) while Equation 5 shows the computation for 23 people (253 potential matches).

$$P(\text{No Matches}) = P(A \neq B) \bullet P(A \neq C) \bullet P(A \neq D) \bullet P(B \neq C) \bullet P(B \neq D) \bullet P(C \neq D) \tag{4}$$
$$= \left(\frac{364}{365}\right) \bullet \left(\frac{364}{365}\right) \bullet \left(\frac{364}{365}\right) \bullet \left(\frac{364}{365}\right) \bullet \left(\frac{364}{365}\right) \bullet \left(\frac{364}{365}\right)$$
$$= \left(\frac{364}{365}\right)^6$$
$$= 0.982$$

Since the probability of success and failure have a sum of one, the probability of finding a birthday match in a group of four randomly chosen people can be computed as 1 – 0.982 = 0.018, or 1.8%. Extending the same logic to a randomly chosen group of 23 people, we can compute that the probability of a birthday match is

$$1 - \left(\frac{364}{365}\right)^{253} = 0.5005, \tag{5}$$

or approximately 50.1%. For 50 people, the probability of a birthday match is

$$1 - \left(\frac{364}{365}\right)^{1225} = 0.9653, \tag{6}$$

or approximately 96.5%.

From these computations, we see that neither the number of potential matches nor the probability distribution follows a linear pattern. Instead, the pattern of potential matches is quadratic while the probability distribution is geometric (Figure 4).

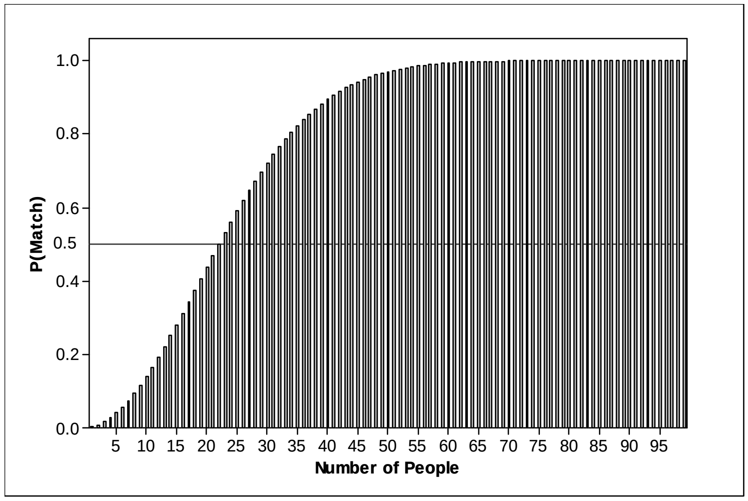


Figure 4. Probability Distribution for a Birthday Match for Groups of 2 – 100 people.

Van Dooren et al. (2003) elaborated on the linearity misconception that leads students to believe that the probability of a birthday match in a group of 23 people is 23/365:

> In that case, you would indeed need 183 people to get a probability for a birthday match exceeding 0.50…We would argue that people applying this strategy would also believe that – compared to a group of 23 – the probability of getting a birthday match in a group of 46 people is doubled, in a group of 69 it is tripled, etcetera (p. 118).

In at least three places throughout the Birthday Problem, intuition typically leads to the application of a linear relationship, sometimes through the modeling of the part/whole relationships that comprise the probability ratio, at others through reverting to linear proportions to analyze the probability space.

These studies seem to indicate that reasoning errors involving rational numbers may be heuristic in nature, but they also may indicate deeply embedded, fundamental misunderstandings about the meaning and relationships of the part whole relationship and how to compare these quantities. Probability, on the other hand, appears to include contextual situations that present students with the opportunity to engage with both the heuristic and meaning difficulties of rational number meaning and the appropriate application of the linear proportion (i.e., discernment between additive and multiplicative relationships) This intertwining of rational number meaning and relationships among rational number quantities with probability contexts can be seen again in the rational number concepts of uniformity, equality, and change

Uniformity, Equality, and Change

Intuitive beliefs about uniformity are highly associated with rational number reasoning. As Fosnot and Dolk (2002) pointed out, wholes must be divided into equal parts in order for a rational number relationship to make sense. Proportionality requires uniformity as well. When this belief is used as a problem solving technique, it is referred to as the *uniformity heuristic* (Falk, 1992, p. 205). In probability, theoretical probabilities are based on the assumption of uniformity, and probability spaces are often assumed to be distributed equally across outcomes. Problems involving conditional probability run counter to uniformity and equality beliefs. Confusion about the meaning of conditional situations and their effect on resultant probabilities (consisting of several rational number quantities) leads to misconceptions about the nature of conditional probability and the effect of increased information on possible outcomes. In the absence of training, students often fall back on the uniformity heuristic, resulting in overgeneralization errors. The Three Prisoner Problem, which is mathematically identical to Vos Savant's (1990) Monty Hall Problem, illustrates the issues surrounding the interpretation of conditional information based on assumptions of uniformity. The problem as described by Bar-Hillel and Falk (1982) and Falk (1992) is as follows:

> Tom, Dick, and Harry are awaiting execution while imprisoned in separate cells in some remote country. The monarch of that country arbitrarily decides to pardon one of the three. The decision who is the lucky one has been determined by a fair draw. He will be freed; but his name is not immediately announced, and the warden is forbidden to inform any of the prisoners of his fate. Dick argues that he already knows that at least one of Tom and Harry must be executed, thus convincing the compassionate warden that by naming one of them he will not be violating his instructions. The warden names Harry. Thereupon Dick cheers up, reasoning: "Before, my chances of a pardon were 1/3; now only Tom and myself are candidates for a pardon, and since we are both equally likely to receive it, my chance of being feed has increased to 1/2" (Falk, 1992, p. 198).

Students often believe, like Dick, that the probability of his being freed has increased from 1/3 to 1/2 because the probabilities must remain equal across the available outcomes. However, such a belief requires the assumption that the choice of a name is randomly chosen, which does not hold in this situation. Instead, the warden, like Monty in the Monty Hall problem, is choosing to disclose one of the outcomes based on information to which he is privy (Falk, 1992). Therefore, the probability of any one of the prisoners being freed is conditional on the warden's information and requires considering the situation from a conditional probability standpoint.

Conditional probability focuses primarily on how quantities represented by rational numbers change as a result of new information being added to a context. The new information alters the assumptions on which the original quantities are based, thereby changing the meaning of the new quantities as well. For such a situation, Bayes' Theorem is especially helpful. Bayes' Theorem defined probability as "the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the value of the thing expected upon its happening" (Bayes, 1763, p. 376). These two values became known as the prior distribution (value based on expectation) and the posterior distribution (value based on an experiment). Although the probability for each of the three prisoners is described as uniform in the problem itself, this description referred only to the prior distribution:

> In the case of an event concerning the probability of which we absolutely know nothing antecedently to any trials made concerning it, seems to appear from the following consideration; viz. that concerning such an event I have no reason to think that, in a certain number of trials, it should rather happen any one possible number of times than another. For, on this account, I may justly reason concerning it as if its probability had been at first unfixed, and then determined in such a manner as to give me no reason to think that in a certain number of trials, it should happen any one possible number of times than another (Bayes, 1763, p. 393).

Uniformity in the prior distribution means that $P(T) = P(D) = P(H) = 1/3$, where the events T, D, and H represent the event of each person being chosen for freedom. Assuming that the guard has no reason to lie and no bias, we can compute the likelihood that he would name either Harry or Tom to not be freed, given that he knows which prisoner will be freed and cannot name Dick (Table 5).

Table 5
Conditional Probability Equations (from Falk, 1992, p. 201)

Probability Equation	Description
P (h | T) = 1	Probability that Harry is named if Tom is to be freed.
P (h | D) = ½	Probability that Harry is named if Dick is to be freed.
P (h | H) = 0	Probability that Harry is named if Harry is to be freed.
P (t | T) = 0	Probability that Tom is named if Tom is to be freed.
P (t | D) = ½	Probability that Tom is named if Dick is to be freed.
P (t | H) = 1	Probability that Tom is named if Harry is to be freed.

Dick arrived at the probabilities in Table 5 under the assumption that he cannot be named. His probabilities also consider the likelihood of each prisoner being freed to be equal. So, if Harry is to be freed, and the warden cannot name the prisoner to be freed, and he cannot name Dick, then only Tom can be named, so P (t | H) = 1. Likewise, if Dick is to be freed, then the warden can name either Tom or Harry, so P (h | D) = P (t | D) = ½. In Bayesian terms, these probabilities represent the prior distribution (in this case, the distribution prior to the warden naming a prisoner). These probabilities, however, do not represent the probability distribution after the warden names the prisoner — the posterior distribution. Bayes' Theorem provides a formula for computing posterior distribution probabilities based on the prior distribution probabilities (Equation 7).

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid \overline{B})P(\overline{B})}, \tag{7}$$

The denominator of this formula represents the total probability of an outcome while the numerator represents the probability of an outcome (Event B) under the condition A. Applying this formula to the prisoner problem, we can compute the posterior probability that Dick will be freed given that Harry was named using Equation 8 (Falk, 1992, p. 201).

$$P(D \mid h) = \frac{P(h \mid D) \bullet P(D)}{P(h \mid D) \bullet P(D) + P(h \mid T) \bullet P(T) + P(h \mid H) \bullet P(H)} \tag{8}$$

The denominator of Equation 8 represents the total probability of any prisoner being freed if Harry is named by the warden. The numerator represents the probability of Dick being freed if the warden names Harry. Substituting the prior distribution probabilities from Table 5 produces Equation 9.

$$P(D \mid h) = \frac{\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)}{\left(\frac{1}{2}\right)\left(\frac{1}{3}\right) + \left(1\right)\left(\frac{1}{3}\right) + \left(0\right)\left(\frac{1}{3}\right)} = \frac{1}{3} \tag{9}$$

So, the disclosure of the warden's information does not change Dick's probability of being freed from the original 1/3. The examination of this outcome brought to light a belief related to the uniformity belief, the *no news, no change* belief (Falk, 1992). While Falk described both of these beliefs as heuristics because subjects used them to solve problems, they are not beliefs about procedures, but of meaning under the context of change. Subjects who believed the no-news-no-change perspective believed that Dick already knew that one of the other two would not be freed, so revealing the name added no new substantive information. Unlike the uniformity heuristic, the no news, no change heuristic correctly computes the solution as 1/3. However, Falk provided two illustrations to demonstrate the erroneous nature of this belief.

First, the no-news-no-change pattern does not hold for Tom. Using Equation 10 and the values from Table 5, we can compute the probability for Tom now that Dick has been told that Harry will not be freed.

$$P(T \mid h) = \frac{P(h \mid T) \bullet P(T)}{P(h \mid D) \bullet P(D) + P(h \mid T) \bullet P(T) + P(h \mid H) \bullet P(H)} \tag{10}$$

Using substitution, we compute Tom's posterior probability (Equation 11).

$$P(T \mid h) = \frac{(1)(\frac{1}{3})}{(\frac{1}{2})(\frac{1}{3}) + (1)(\frac{1}{3}) + (0)(\frac{1}{3})} = \frac{2}{3} \tag{11}$$

So, although Dick's chances did not change by the warden's information, Tom's chances of being freed have doubled — the warden's information was not completely irrelevant to the probabilities. So, if Dick could choose, switching places with Tom would double his chances of being freed. This outcome is isomorphic with the Monty Hall problem: Switching doors also doubles a contestant's chances of winning the prize. The Monty Hall problem attracted a great deal of attention in the late 1980's and early 1990's as mathematicians and mathematics educators vehemently opposed Vos Savant's (1990) claim that the choice of switching doors made a difference in the probability of winning. This opposition to the switching claim adds further evidence to Van Dooren et al.'s (2003) conclusions about the pervasiveness and persistence of mathematical misunderstandings about meaning and relationships: Such errors are evidently not limited to novice learners.

Falk's second argument against the no-news-no-change belief related the belief back to uniformity; specifically, she showed that this belief also relied on an assumption of uniformity. Falk (1992) related a variation of the prisoner problem in which Tom is favored by the monarch, and so he gets two votes for freedom while Dick and Harry each receive only one. The prior probabilities become $P(T) = 1/2$, $P(H) = P(D) = 1/4$, a non-uniform distribution. The assumptions about the decision making process for the warden remain unchanged (i.e., no bias or reason to lie). Using Bayes' Theorem again, the posterior probability for Dick getting his freedom changes, as shown in Equation 12.

$$P(D \mid h) = \frac{\left(\frac{1}{2}\right)\left(\frac{1}{4}\right)}{\left(\frac{1}{2}\right)\left(\frac{1}{4}\right) + \left(1\right)\left(\frac{1}{2}\right) + \left(0\right)\left(\frac{1}{4}\right)} = \frac{1}{5} \tag{12}$$

When the assumption of uniformity does not hold true for a situation, assumptions about change and the meaning of rational number quantities under conditions of change become untenable. The ability to discern the applicability of any particular assumption about rational number quantities for a specific problem requires understanding of which units to count, how to count them, and how those quantities interact with other information from a particular problem (i.e., relational understanding, as in Skemp, 1976/2006). Freudenthal (1970) summarized the unique role of probability in connecting these abstract mathematical concepts to concrete examples:

> Probability applies in everyday situations…There is no part of mathematics that is as universally applied except, of course, elementary arithmetic…In no mathematical domain is blind faith in techniques more often denounced than in probability; in no domain is critical thought more often required (p. 167).

These studies suggest that probability instruction may play a unique role in challenging deeply-held assumptions about the meaning and relationships of rational number quantities within a contextual situation.

Critical Misconceptions Specific to Learning Algebra

Students beginning the study of algebra face learning barriers from several sources. First, algebra is often the first course in which students are asked to engage in abstract reasoning and problem solving (Vogel, 2008). Researchers have demonstrated that the abstract nature of algebra increases its difficulty over arithmetic (Carraher &

Schliemann, 2007; Howe, 2005; Kieran, 1989). The impediment of abstractness to the construction of meaning directly affects the ability of students to construct multiple representations of algebraic objects (Kieran, 1992; Vogel, 2008).

Second, the learning of algebra requires students to learn a language of mathematical symbols that is completely foreign to their previous experiences (Kilpatrick et al., 2001). The multiple ways in which this language is described and used during instruction often prevents students from connecting algebraic symbols to their intended meaning (Blanco & Garrote, 2007; Socas Robayna, 1997). In some cases, students are completely unaware that any meaning was intended for the symbols (Küchemann, 1978). In other cases, they may know that meaning exists, but limited understanding prevents them from ascribing meaning to the symbols, or they may assign erroneous meaning to the symbols (Küchemann, 1978). For example, as students study topics such as functions and graphs, they begin to understand and interpret one set of algebraic objects in terms of another (e.g., a function equation with its graph, a data set by its equation, a data set by its graph, as in Leinhardt, Zaslavski, & Stein, 1990). McDermott, Rosenquist, and Van Zee (1987) found that students are generally able to plot points and equations; however, in spite of this procedural fluency, students still lack the ability to extract meaning from graphical representations. They concluded that the difficulty lay in the connection of a graph to the construct it represents. Specifically, students are readily capable of demonstrating procedural fluency, but memory and procedural understanding is unable to guide students through problems involving interpretation (Skemp, 1976/2006).

Kieran (1992), Howe (2005), and Carraher and Schliemann (2007) recognized that learning the structural characteristics of algebra creates a third obstacle faced by students.

> The difficulty that students experience with understanding the structure of algebra, even its most elementary aspects such as are found in high school textbooks, was exemplified by their early attempts to convert expressions into equations in order to have a representation that includes a result, the unsystematic and strategic errors they committed while simplifying expressions, their resistance to operating on an equation as an object as shown by their not using the solving procedure of "doing the same thing to both sides," their not treating the equal sign as a symbol of symmetry…their difficulty in seeing the "hidden" structure of equations, [and] their non-use of algebra as a tool for proving numerical relations (Kieran, 1992, p. 412).

The abstract, structural, and language barriers interact within algebra. For example, consider the expression $a + b$: How students interpret the meaning of each variable depends on how well they can handle the abstract nature of the symbols. Further, students must recognize that the expression $a + b$ represents the total number of items from a set of a and b items (Kieran, 1992).

The teaching methods used to convey content often create a fourth barrier to learning algebra. Sfard (1991) highlighted a difficulty of expectation as one problem with teaching methods:

> More often than not, both students and teachers fail to acknowledge the fact which is one of the most important implications of our three-phase schema: Insight cannot always be expected as an

immediate reward for a person's direct attempts to fathom a new idea. The reification, which brings relational understanding, is difficult to achieve, it requires much effort, and it may come when least expected … [it] may occur after a period of intensive work followed by days of rest (p. 33).

Kieran (1992) concluded from Sfard's (1991) study that a great deal of time must be spent connecting algebra to arithmetic before proceeding to the structural ideas of algebra. Furthermore, the lack of materials designed to facilitate the transition from arithmetic to algebra forces teachers to either create materials themselves or conduct time-intensive searches (Kieran, 1992). Instead, teachers often rely on whatever sequence is outlined by a textbook.

Such a choice is highly problematic: Kieran (1992) proposed that textbook explanations are often insufficient for helping students understand the abstract, structural concepts necessary in algebra. Consider the following explanation of linear functions from an algebra textbook (Figure 5):

Graph the equation $y = -2x - 1$.

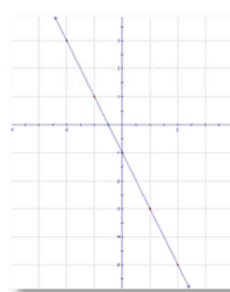
Solution

STEP 1 Construct a table of values.

x	-2	-1	0	1	2
y	3	1	-1	-3	-5

STEP 2 Plot the points. Notice that they all lie on a line.

STEP 3 Connect the points with a line.



LINEAR FUNCTIONS The function $y = -2x - 1$ in Example 4 is a **linear function** because it can be written in the form $y = mx + b$ where m and b are constants. The graph of a linear function is a line. By renaming y as $f(x)$, you can write $y = mx + b$ using **function notation**.

$y = mx + b$ Linear function in x-y notation

$f(x) = mx + b$ Linear function in function notation

Figure 5. Textbook description of linear functions (Larson, Boswell, Kanold, & Stiff, 2010, p. 75)

Function Relationships

Notice that the definition of linear functions as described in this example relies on recognizing a prescription: If a function can be written in slope intercept form, it is linear. While not incorrect, this explanation is insufficient for situations that call for alternate forms of linear functions (such as the standard form or point-slope form). Notice also that the definition of function notation is mixed with the definition for linear functions. Students often fail to recognize that function notation is a general form intended for all functions rather than just linear (Chang, 2002). As a result, students may develop a misconception that functions are supposed to be linear (Chang, 2002; Kalchman & Koedinger, 2005). Socas Robayna (1997) offered another example: Students may continue trying to simplify an expression until they reduce it to a single number. Baroudi (2006) noted similar difficulties with the meaning of the equal sign, and found that additional time spent with numerical equations may not be sufficient for learning the structure of algebraic equations. Instead, he suggested the importance of intermediate representations to bridge the gap between arithmetic and algebraic structures.

Skemp (1976/2006) considered the underlying foundations of mathematical misconceptions as emerging from an instrumental understanding of mathematics that forces students to rely on memorization. Kieran (2007) agreed with Skemp's viewpoint of the limiting nature of instrumental mathematics. Even the manipulation of symbols, once considered primarily an algorithmic process, has become recognized as emergent from concepts (Kieran, 2007). Skemp gave the analogy of a person trying to navigate through a new city. A person with an instrumental understanding of the city may have a number of ways to get from point A to point B. The difficulty with this understanding arises when the person deviates from the original course. In such a case, the person gets lost. Instrumental understanding of algebra produces similar results. For instance, students may learn a set of prescriptions for solving equations of the form $ax + b = c$; when they encounter equations of the form $ax + b = cx + d$, their prescriptions are unable to accommodate the new form.

Probability and rational number assumptions influence how students understand non-linear functions. Student understanding of rational number and probability concepts may also influence their understanding of algebraic structures (Falk, 1992).A famous examples of this misuse of the rule of three took place as a result of a bad bet:

> De Méré knew that it was advantageous to bet on the occurrence of at least one six in a series of four tosses of a die – maybe this was an old experience. He argued it must be as advantageous to bet on the occurrence of at least one double-six in a 24 toss series with a pair of dice. As Fortune disappointed him, he complained to his friend Pascal about preposterous mathematics which had deceived him (Freudenthal, 1970, p. 151).

De Méré made two erroneous assumptions about the probability and rational number structures in this situation. First, he assumed the one die probability to be 4/6, computed by adding the probability of 1/6 four times (i.e., an additive, linear accumulation of probabilities). Second, he inferred that the probability of rolling at least a double six with two dice should be proportional, or 24/36. Both situations are binomial rather than linear. For the one-die scenario, P(Success) = 1/6 while P(Failure) = 5/6. The binomial theorem where $n = 4$ and $k \geq 1$yields Equation 13.

$$P(X \geq 1) = \binom{4}{4}\left(\frac{1}{6}\right)^4 + \binom{4}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right) + \binom{4}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2 + \binom{4}{1}\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^3 = \frac{29}{56} \text{ or } 0.517 \qquad (13)$$

So, rather than having a 67% probability, the original probability should have been 51.7%. In the two-dice scenario, P(Success) = 1/36 while P(Failure) = 35/36. The binomial theorem where $n = 36$ and $k \geq 1$ can be computed similarly to produce the probability of rolling at least one double six in 24 rolls to be 0.491 or 49.1%. The binomial formula, a non-linear algebraic equation, applies to many situations where linear relationships do not hold for the quantities of interest. The De Méré problem illustrates how these underlying assumptions about probability and rational number structures influence understanding of the functional relationship that is so critical to algebraic thinking and reasoning when that relationship is not linear (Freudenthal, 1983; Kalchman & Koedinger, 2005; Kaput & Hegedus, 2004; Thorpe, 1989).

Probability and rational number assumptions influence how students understand linear functions. Even algebraic problems that do require a linear function cause students tremendous difficulties (Moss, Beatty, Barkin, & Shillolo, 2008). Moss et al. attributed these difficulties to student misconceptions about additive versus multiplicative structures. For example, the Trapezoid Table Problem presents a series of trapezoidal tables joined with seats placed around the table. Students are also provided with a table of values as in Figure 6.

Table	Chairs
1	5
2	8
3	11
n	

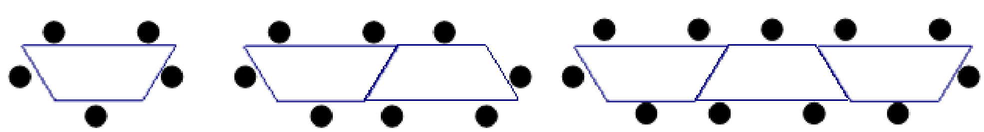


Figure 6. Representations provided in the Trapezoid Table problem (Moss et al., 2008, p. 157).

For pattern problems such as the Trapezoid Table, Warren (2000) found that students struggling with linear patterns tended to revert to a recursive pattern (i.e., repeatedly adding three to the previous y value, $y_n = y_{n-1} + 3$) rather than developing a functional relationship between the number of tables and number of chairs (i.e., recognizing that the three acts as a slope ratio of change, a multiplicative operand, $y_n = 3n + 2$). Stacey (1989) identified the same tendency with similar problems (e.g., the Christmas Tree Problem, Ladders). The problem with such thinking is not that the recursive pattern is incorrect, but that it does not describe the relationship between x and y as students often believe it should (Warren, 2000).

Non-linear functions offer a similar challenge for algebra students; specifically, students continue to apply this additive feature to numerical relationships, confusing it for functionality. If the additive feature is present, as in the Trapezoid Table problem, then students recognize the sequence as a function. If, on the other hand, an additive pattern cannot be found, the relationship is discarded as non-functional (Chang, 2002; Kalchman & Koedinger, 2005). More generally, Clement (2001) noted that students often rely on the presence of a formula do determine if a relationship is a function:

> Students may erroneously consider $\pm\sqrt{x^2 - 3}$ a function, since it is an algebraic formula; whereas they might not consider the correspondence that Mary owes \$6, John owes \$3, and Sue owes \$2 to be a function, since no formula "fits it" (p. 746).

Variables and Variation

Variable interpretation. The notion of variability is especially important in algebra (Briggs, Demana, & Osborne, 1986; Edwards, 2000; Graham & Thomas, 2000;

Kalchman & Koedinger, 2005). MacGregor and Stacey (1997) found that students have difficulty assigning meaning to variables, failing to recognize the systemic consistency in the multiple uses of variables. Research efforts in algebra have long focused on how well students could discriminate between the uses of variables (Kieran, 2008). Küchemann (1978) developed a test for variable understanding which matched Piagetian sub-stages with item complexity: The results indicated that students interpret variables six ways (Table 6).

Table 6
Hierarchical Levels of Variable Interpretation

Level	Piagetian Sub-Scale	Description
1	Concrete Operations	Evaluating the variable using trial and error.
2		Ignoring the variable.
3		Variable represents an object or label.
4	Formal Operations	Variable represents a specific unknown.
5		Variable represents a generalized number.
6		Variable represents a functional relationship.

Gray, Loud, and Sokolowski (2005) examined student responses to questions examining student interpretation of variables using Küchemann's hierarchy as a framework. For the question, "Small apples cost 8 cents each and small pears cost 6 cents each. If *a* stands for the number apples bought and *p* stands for the number of pears bought, what does 8a + 6p stand for?" 81% of students in basic algebra, 76% in college algebra, and 50% in calculus answered incorrectly (Gray et al., 2005, p. 4). Gray et al. identified the most common error as substituting the price of the fruit for the letters and giving the resultant solution, 100. Students who gave this solution appeared to interpret *a* and *p* as specific unknowns (Küchemann's Level 4). The next common error resulted from interpreting the letters as labels for the objects (Küchemann's Level 3) rather than the price of the objects, "8 apples and 6 pears" (p. 5). Their findings agreed with those found by Küchemann (1981) for high school students: "Children have relatively little difficulty with items…where the letters can be thought of as objects or names of objects…they find it much more difficult when the letters necessarily represent numbers, especially numbers of objects" (p. 307).

Similarly, Torigoe and Gladding (2006) compared student ability to solve sets of parallel problems, one involving numerical values and the other providing variables. In one problem, students were asked to determine the minimum acceleration necessary for police to catch a bank robber fleeing the scene of a crime. In one version, the prompt provides specific numerical quantities while the second version provided symbols to represent the quantities.

The percentage of correct responses for the symbolic version (57%) was significantly lower than for the numerical version (94%) for a sample of 894 college students. Their study suggested that, holding all other task characteristics constant, the meaning of variables and the quantities they represent causes significant difficulties for students in algebra. Thorpe (1989) suggested that one possible reason for such difficulty is in the fragmentation of instruction. He encouraged the elimination of the concept of expressions from the algebra curriculum entirely:

> Asking students in an algebra course to manipulate expressions is analogous to asking students in a writing course to manipulate phrases rather than sentences. Expressions are not important in themselves. They are important only when they are implicitly or explicitly part of an equation. The expression $2x + 1$, by itself, is incomplete. To have meaning, it must be imbedded in an equation, such as $f(x) = 2x + 1$, or $2x +$

> 1 = 0. The equation provides meaning for the expression, as well as a context for *x*. (Is *x* a variable or does *x* represent a member of a solution set?) Just as we teach students of writing to speak in sentences, let us teach students of algebra to speak in sentences! (p. 18).

From Thorpe's point of view, understanding the nature of a variable is intertwined with the meaning of equations in a particular context. However, placing variables in the context of an equation may not be sufficient to advance student understanding of variables.

Operating at Küchemann's (1978) Level 4, many students can readily find the solution of 3 for a problem such as $5x - 4 = 11$. However, in solving for such an unknown, students may not recognize the varying nature of *x*; that is, that as *x* takes on different values, the value on the right hand side of the equation (i.e., the 11) changes as well. Furthermore, the same letter may be used in multiple problems. Suppose instead that $5x - 4 = 20$. In this problem, we find that the same letter *x* now represents a value of 3.2. Students rarely recognize or value such subtleties of change, instead relying on rote procedures (Fuson et al., 2005; Kalchman & Koedinger, 2005), nor do they recognize the connection to a two variable equation, such as $y = 5x - 4$ (Kieran, 2008). Students will often look for an "answer," not recognizing that multiple solution sets can exist within a single problem or that multiple equations can be related. This lack of meaning may also be due to fragmented instruction (Thorpe, 1989; Kieran, 1989, 2007) and an instructional focus on procedures rather than concepts and connections between ideas (Hiebert & Carpenter, 1992; Hiebert & Grouws, 2007).Probability instruction can be used to focus the concept of variable directly on the changing values within a quantity, thereby helping students avoid or alleviate this confusion.

The reversal error presented by Kaput and Clement's (1979) Student Professor Problem (i.e., Write an equation to represent the phrase, 'There are six times as many students as professors at this university') may demonstrate underlying misconceptions in the meaning of variables and equality. Clement (1982) attempted to eliminate the reversal error by warning students of the potential reversal while Rosnick and Clement (1980) tutored students specifically about the reversal error, hoping that cognitive awareness alone could assist students. Fisher (1988) and Phillippe (1992) substituted used letters other than *S* and *P*, hoping to advance students beyond the use of letters as labels. Each of these efforts resulted in a lack of significant change in the error rate. Clement (1982) identified three types of strategies used by students in the student professor problem (see Figure 7).

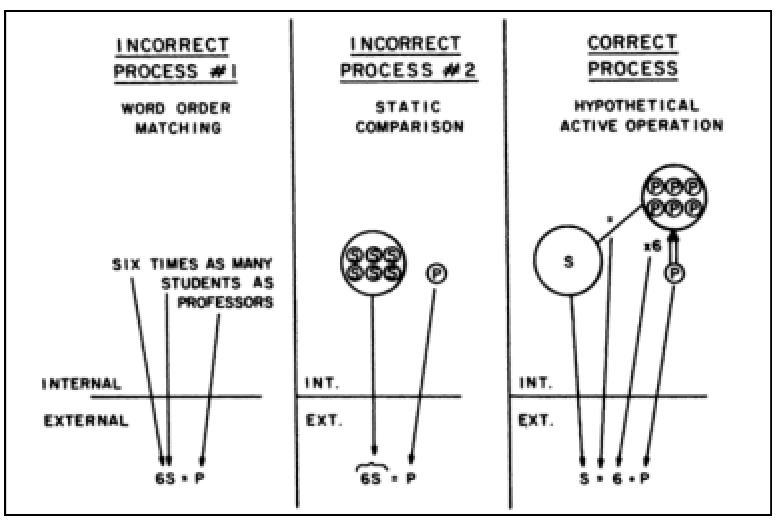


Figure 7. Strategies Used in the Student Professor Problem (Clement, 1982, p. 21).

In the word order matching strategy, students displayed the label misconception, operating at Küchemann's (1978) Level 1. Similarly, in the static comparison, students relied on a mental image of the makeup of a typical university with more students than professors. Clement (1982) found that the static comparison image was so strong that even after considering the correct equation, $S = 6P$, students considered it impossible and discarded it in favor of the erroneous equation.

Probability quantities are also affected by static comparisons. Static comparison thinking may contribute to misconceptions in probability as well. Consider the following problem: "Of the following two subgroups, which is larger? (a) Unmarried physicians, or (b) Unmarried physicians who like to travel abroad" (Agnoli & Krantz, 1989, p. 543).

Respondents overwhelmingly chose (b), reasoning that unmarried physicians are more likely to travel abroad than married ones. This reasoning also represents static comparison: Agnoli and Krantz (1989) found that students based their decisions on a priori knowledge rather than the meaning of the mathematical statement. Likewise, MacGregor and Stacey (1997) suggested that students often make such errors as a result of relying on intuition and making analogies to more familiar situations. Errors resulting from static comparisons are persistent, resistant to direct interventions, and often result in biased mathematical judgments in unfamiliar contexts (Shaughnessy & Bergman, 1993).

Kahneman and Tversky (1972, 1973a, 1973b, 1982, 1983), Agnoli and Krantz (1989), and Shaughnessy and Bergman (1993) divided these biased judgments into two categories, representativeness and availability. Representativeness, results from transferring properties of large samples to small samples, especially the notion that small samples reflect the parent population as well as large samples (or, "neglect of small samples" in Shaughnessy & Bergman, 1993, p. 182). The second judgmental heuristic error, availability, reflects a person's tendency to estimate probabilities based on available personal experiences. These judgment errors themselves do not appear to be misconceptions, but they do appear to emerge from a fundamental misunderstanding about the nature of probabilistic quantities. Whether static comparison errors begin with probability or algebra or affect both content domains simultaneously is unclear. These errors may, however, link fundamental probability misconceptions to algebra misconceptions.

MacGregor and Stacey (1997) suggested that misleading teaching practices and materials may exacerbate the need for students to rely on judgmental heuristics in mathematics. Thorpe (1989), Kieran (1989, 1992) and Leitzel (1989) suggested that de-contextualizing the algebra curriculum may be one such misleading practice that guides students away from the systemic structure of algebra. On the other hand, Kieran (2008) noted that advances in instructional technology may help reverse some of this de-contextualization. For example, Ainley, Bills, and Wilson (2004) presented evidence that spreadsheet applications can help students build bridges from numerical specificity to variable generality:

> In the algebra-like notation of the spreadsheet, the cell reference is used ambiguously to name both the physical location of a cell in a column and row, and the information the cell may contain. The spreadsheet thus offers a strong visual image of the cell as a container in two which numbers can be placed…The image offered by the spreadsheet is ambiguous in another powerful way: when a formula is entered in a cell, it can be 'filled down' to operate on a range of cells in a column. The cell reference can then be seen as both specific (a particular number I may put in this cell) and general (all the values I may enter in this column). This

> image is likely to support the idea of variable as a range of numbers in functional relationships (p. 2).

Chazan and Yerushalmy (2003) examined the complexity of variable concepts and concluded that a functions-based approach focusing on variation/change rather than unknowns to be solved allows students to develop more advanced understanding of variables. Such an approach directs students to interpret variables at Küchemann's (1978) two highest abstract levels, generalized numbers and functional relationships. The concept of variation as studied in probability follows Chazan and Yerushalmy's advice: Variables are quantities within which patterns for expected values are based on distributions; Variation is examined as a measure of change that describes differences between small samples and the population, variation from the mean within a sample; and students explore patterns within random variables and the significance of small amounts of variation in large samples (Watson & Kelly, 2005; Watson, Kelly, Callingham, & Shaughnessy, 2003; Watson & Shaughnessy, 2004; Zawojewski & Shaughnessey, 2000).

Probability instruction, therefore, may offer an alternative approach to leading students to the meaning of variable by combining the concept of variable with authentic contexts and technological tools (e.g., spreadsheets) with the exploration of variation for different sample sizes. Often, students underestimate the amount of variation in small samples, inappropriately applying the Law of Large Numbers (Shaughnessy, 1992; Shaughnessy & Bergman, 1993; Shaughnessy, Canada, & Ciancetta, 2003). Consider the task depicted in Figure 8 (used in Watson & Shaughnessy, 2004: NAEP (1996) Released Item).

Suppose we have a container with 100 candies in it: 50 are red, 20 are yellow, and 30 are green. The candies are all mixed up in the container.

(1A) Suppose you pick out 10 candies.

(i) How many reds do you expect to get?__________
Suppose you did this several times [replacing previously drawn handfuls]. Do you think this many would come out every time?
Why do you think this?

(ii) How many reds would surprise you?__________
Why do you think this?

Figure 8. The Candy Problem (Watson and Shaughnessy, 2004, p. 107).

Student responses on this item indicated that some students have no intuitive sense of the amount of variation they should expect in repeated sampling with replacement while others included a reasonable amount of variation in their predictions (Watson & Shaughnessy, 2004).

Green's (1982) thumbtack question demonstrated how issues with variation are linked with assumptions of uniformity and equality. In this problem, students were asked, "A packet of 100 drawing pins is emptied out onto a table. Some drawing pins land pointing up and some land pointing down: How many up and how many down would you expect out of the 100?" (p. 30). Green (1983a) reported that most students chose a 50-50 outcome, assuming that the probability of up and down is equal. Trying a variation of the same problem, he included additional information about a prior trial in which 32 tacks landed up and 68 down. Some students chose a reversed solution, 64 landed up and 36 down, explaining that the given information did not match their own experiences (i.e., the availability heuristic). The item was modified to its final form, making the prior trial seem more realistic and including a non-numeric choice that all outcomes are equally likely. In the final sample, students overwhelmingly chose the non-numeric option, confirming their belief in uniformity and equality.

Another question from Green (1982) linked notions of variation with randomness and the belief that randomness means uniformity:

> A teacher asked Clare and Susan each to toss a coin a large number of times and to record every time whether the coin landed Heads or Tails. For each 'Heads,' a 1 is recorded and for each 'Tails,' a 0 is recorded. Here are the two sets of results:
>
> | | |
> |---|---|
> | CLARE | 01011001100101011011010001110001101101010110010001
01010011100110101100101100101100100101110110011011
01010010110010101100010011010110011101110101100011 |
> | SUSAN | 10011101111010011100100111001000111011111101010101
11100000001000101001000000100011000101000000000011001
00000001111100001101010010010011111101001100011000 |
>
> Now one girl did it properly, by tossing the coin. The other girl cheated and just made it up. Which girl cheated? How can you tell? (p. 27).

Students at all grade levels overwhelmingly believed that the regularity of Clare's pattern and the long run lengths in Susan's pattern made Susan the most likely culprit for having cheated, when, in fact, the reverse was true. Green conjectured that their reasoning errors emerged from a deeper misconception about the nature of variation within randomness and suggested a link between variation misunderstandings and beliefs about uniformity.

The gambler's fallacy is another example of a reasoning error that may be linked to misconceptions about variation, representativeness, and randomness (Falk & Konold, 1994; Tversky & Kahneman, 1971). The gambler's fallacy denotes a belief that, given a sequence of independent events repeated a number of times and a particular outcome has occurred more than would normally be expected, a different outcome is more likely on subsequent trials. For example, a student tosses a coin six times and gets a single tail the first time and then five heads. The gambler's fallacy represents the belief that the tails outcome is more likely on the next flip.

> When subjects are instructed to generate a random sequence of hypothetical tosses of a fair coin…they produced sequences where the proportion of heads in any short segment stays far closer to .50 than the laws of chance would predict…Subjects act as if every segment of the random sequence has strayed from the population proportion, a corrective bias in the other direction is expected (Tversky & Kahneman, 1971, p. 106).

Tying Algebra, Probability, and Rational Numbers Together through Error Patterns

Static comparisons and judgment bias errors appear to affect the learning of both algebra and probability. Some of these errors may be due to misunderstanding fundamental concepts in algebra such as variable meaning and functions. Errors in reasoning within probability may sometimes be due to misconceptions about rational number quantities and their relationships within specific contexts. Misapplication of additive and multiplicative structures in algebraic contexts may also be connected to misunderstanding the rational number quantities within a specific situation. The connections between these reasoning errors also appear in geometry contexts.

Critical Misconceptions Specific to Learning Geometry

Spatial Reasoning

Student orientation toward geometry is quite different from that of algebra: Students are often intrinsically motivated to study the properties that govern the shapes encountered in daily life (Engel, 1970; Freudenthal, 1973). In spite of this motivational factor, students still struggle with errors and misunderstandings in geometry due to limited spatial reasoning (Clement & Battista, 1992).

Spatial reasoning begins with the differentiation between objects and representations. Objects are abstractions, ideas considered through reasoning (Battista, 2007). Representations, on the other hand, are used to signify objects other than themselves. For example, a line drawn on a piece of paper only represents a geometric line, defined in Euclidean geometry as having infinite length and no thickness. The Van Hiele (1959/1984a) framework is especially helpful for describing student spatial reasoning processes and how they distinguish between objects and representations.

This framework classifies geometric reasoning into five levels. Within Level 0, the base level (*visual* in Clement & Battista, 1992), children reason geometrically solely on the basis of recognition. At this stage, shapes are examined as a whole. Only the physical appearance of a shape is considered without regard to parts or properties, and no distinction is made between objects and representations (Crowley, 1987). For example, they may recognize that a rectangle is different than a square only because it appears different. Within Level 1 (*analysis* in Crowley, 1987 and *descriptive/analytical* in Clements & Battista, 1992), students analyze geometric concepts using properties and characteristics of shapes and figures (Crowley, 1987). For example, students will recognize that a rectangle has four sides, that opposite sides are equal, and that the four angles all measure 90°. They may not, however, recognize the hierarchical ordering of properties. For example, they may attribute the properties of a rectangle to a square without realizing that a square is actually a special rectangle. At Level 2 (*abstract/relational*; Clements & Battista, 1992), students do begin the ordering of properties (Van Hiele, 1959/1984a) through *informal deduction* (Crowley, 1987). Students using abstract/relational reasoning categorize shapes and figures according to their properties and recognize hierarchical classifications, such as considering a square to be a rectangle with congruent sides. At Level 3 (*deduction* in Crowley, 1987 and *formal deduction* in Clements & Battista, 1992), students are able to develop theorems within an axiomatic system. Additionally, they distinguish between the roles of theorems, postulates, as definitions; their thinking is also concerned with the meaning of the converse of a theorem (Crowley, 1987; Van Hiele, 1959/1984a). Most high school instruction goes no further than Level 3 (Crowley, 1987); however, Level 4 (*rigor* in Crowley, 1987 and *rigor/mathematical* in Clements & Battista, 1992), involves the formal reasoning about mathematical systems in the absence of reference models. Clements & Battista (1992) also proposed a pre-base level which they called *pre-recognition* in which children are unable to distinguish between shapes.

The Van Hiele framework can be used to help explain common misconceptions that develop in geometry through missing or inadequate spatial reasoning. Clements and Battista (1992, p. 422) compiled 11 of the most common geometric misconceptions:

1. An angle must have one horizontal ray.
2. A right angle is an angle that points to the right.

3. To be a side of a figure a segment must be vertical.
4. A segment is not a diagonal if it is vertical or horizontal
5. A square is not a square if its base is not horizontal.
6. The only way a figure can be a triangle is if it is equilateral.
7. The height of a triangle or parallelogram is a side adjacent to the base.
8. The angle sum of a quadrilateral is the same as its area.
9. The Pythagorean Theorem can be used to calculate the area of a rectangle.
10. If a shape has four sides, then it is a square.
11. The area of a quadrilateral can be obtained by transforming it into a rectangle with the same perimeter.

Students who hold Misconceptions 1 – 5 operate at the base level of recognition. For example, a student who believes that a square is not a square unless its base is horizontal (Misconception 5) does not associate the properties of a square to the label. Instead, such a student relies strictly on the visual orientation of a particular drawing. Students who hold Misconceptions 6 – 11 have moved to the analysis level: They are aware of properties, but the properties have not been organized into a coherent system. For example, a student who believes that the sum of a quadrilateral is the same as its area (Misconception 8) acknowledges that a quadrilateral has the property of a constant sum for its interior angles but confuses the meaning of an angle with the meaning of area.

Teaching methods and materials coupled with a lack of authentic experiences may exacerbate misconceptions resulting from the limited spatial reasoning found at the lower Van Hiele levels (Oberdorf & Taylor-Cox, 1999). They described the example of early geometric activities in which students are taught to distinguish between rectangles and squares. They maintained that such differentiation is quite difficult to eradicate in later grades. Instead, they advocated for the use of exploratory activities that allow students to examine quadrilaterals as a whole and provide the substance for rich discussions about similarities and differences between different quadrilaterals.

Monoghan (2000) found that textbooks tend to reinforce an over-reliance on typical representations of geometric objects, a condition that may result in limiting progression from the recognition stage to the analysis stage. Swindal (2000) and Monaghan (2000) recognized a fundamental gap first identified by Van Hiele (1959/1984a) and Shaughnessy and Burger (1985): Students and teachers think about the same concepts from different levels. Most students in high school geometry reason at Levels 0 or 1 (recognition and analysis) while teachers think, reason, and teach using vocabulary from Level 2 (abstract/relational thinking). Furthermore, courses that focus primarily on the development of proof using language from Level 3 offer most students, who are functioning at Levels 0 and 1, limited opportunity to advance their understanding of spatial properties and relationships (Hiebert & Grouws, 2007).

Misunderstandings about spatial properties and relationships appear dissimilar to algebra, probability, and rational number reasoning errors on the surface, but the number patterns within spatial relationships involve rational numbers and algebraic patterns. Geometric models are also often used to represent probability ratios.

Proportionality and Geometric Learning

Just as with rational numbers, probability, and algebra, linear proportions abound in geometry (e.g., side lengths and perimeter of similar figures follow the Rule of Three, as in Carter et al., 2010; Dietiker et al., 2007; Serra, 2003). Freudenthal (1983) and Stacey (1983) noted that the abundance of linear applications in geometry often leads students to the belief that linearity is universally applicable. Students often cling to the linear model tenaciously in spite of additional information that discredits the linear model for a particular scenario (De Bock et al., 1998, 2002; De Bock, Van Dooren, Verschaffel, & Janssens, 2002). De Bock et al. (1998, 2002) studied student problem solving with problems involving squares, circles, and irregular figures, half of which required a linear proportion while the other half required non-linear reasoning. Their example for square figures follows (De Bock et al., 1998, p. 68).

Enlargement of a square figure

Proportional item:

> Farmer Gus needs approximately 4 days to dig a ditch around a square pasture with a side of 100 m. How many days would he need to dig a ditch around a square pasture with a side of 300 m? (Answer: 12 days)

Non-proportional item:

> Farmer Carl needs approximately 8 hours to manure a square piece of land with a side of 200 m. How many hours would he need to manure a square piece of land with a side of 600 m? (Answer: 72 hours)

De Bock et al. (1998) found that 98% of their sample of 12 and13 year old students solved the proportional problems correctly, whereas only 5% of the same sample solved the non-proportional items correctly. They also found that problems for irregular figures were missed more than problems for squares or circles. In their follow up study (De Bock et al., 2002), they interviewed students who had missed non-proportional problems. The interviews progressed through five stages, each progressively adding more information to direct the students toward a non-linear model. Most students required at least three stages before realizing that the linear solution was incorrect; some students clung to the linear solution even after all five stages.

The ability to distinguish how and when to use proportionality relationships appears to affect the learning of geometry as well as algebra, probability, and rational numbers. Misusing these relationships may be due to misunderstanding geometric ideas and the connections between them. The inter-connectedness between these potential misunderstandings with rational number, probability, and algebra may indicate that a novel teaching strategy targeting the underlying misconceptions may help reduce reasoning errors in all four content areas.

Teaching Probability to Correct Foundational Mathematical Misunderstandings

Stone, Alfeld, and Pearson (2008) echoed the sentiments of Freudenthal (1970): In order to guide students to deep mathematical learning, mathematical content must be tied to authentic experiences to which students can relate. Probability offers such a connection between mathematics and the real world naturally (Liu & Thompson, 2007), and its de-emphasis in U.S. high school mathematics curricula may account for many of the difficulties students have connecting abstract mathematical ideas to concrete examples (Davis, 1992). In spite of the ability of probability to bridge the gulf between the abstract and concrete, several reasons explain its exclusion from mathematics curricula. First, teachers are typically less familiar with probability content than other areas of mathematics (Jendraszek, 2008; Swenson, 1998). Compounding this problem is the fact that probability is often viewed as a second-rate topic (Mitchell, 1990;

Shaughnessy, 2006). Furthermore, curriculum issues in the United States have historically been problematic: Every state develops its own standards, varying widely in organization and complexity (Boland & Nicholson, 1996; Reys & Lappan, 2007). Issues of cognitive development of a child and student mobility between schools and states compound curriculum issues even further (Engec, 2006; Fajemidagba, 1983). The National Council of Teachers of Mathematics (NCTM) began an effort to coordinate the development of a recommended mathematics curriculum, publishing the *Curriculum and Evaluation Standards* as a result (NCTM, 1989). Even after concerted efforts to increase the teaching of probability, Shaughnessy (1992) found that the NCTM recommendations are minimized in the classroom. Figure 10 shows that, even when the recommendations are followed, number, algebra, and geometry receive the greatest emphasis while probability is given minimal attention.

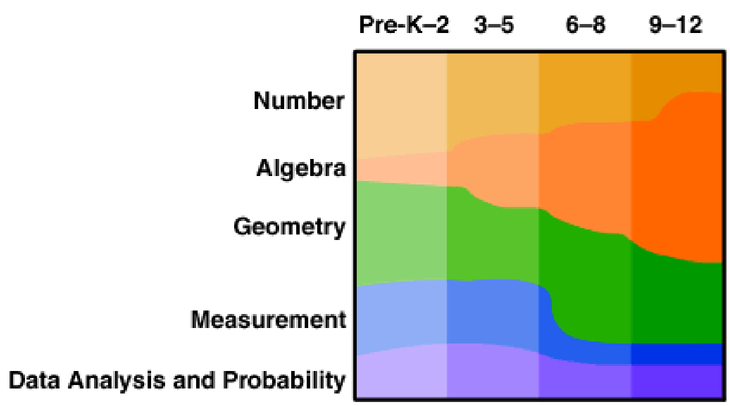


Figure 9. NCTM (2000) Mathematics Strands for Grades K – 12 (p. 30).

Although data analysis and probability are inextricably linked, the two areas may need separate degrees of emphasis in high school. Shaughnessy (2007) found that student ability to compute means, medians, and modes had improved since his 1992 report, but student understanding of randomness, chance, and variation had not improved correspondingly. Likewise, Smith (2003) found that many high school teachers handle probability differently than statistics; they may relegate probability to the end of the year as time permits or simply delete it from the curriculum completely.

This de-emphasis of probability has devastating consequences for mathematics students. First, excluding probability from the mathematics curriculum may increase the disparity between mathematics and the real world. Shaughnessy and Bergman (1993) stated, "It appears that stochastic problems may closely resemble the type of problem solving that our students will have to do in their own private lives or on their jobs" (p. 193). Furthermore, people are faced with choices involving probability on a daily basis that affect the quality of life for themselves and their family and friends, e.g., career decisions (Hume, 1970; Papps, 2008), interpreting weather, economic, and political forecasts (Resnick, 1987), business and personal purchasing choices (Ashman, 2001; McAvoy, 2001; Swaminathan, 2003), and gaming (Barry, 1988; Brandt & Pietras, 2008; Clotfelter & Cook, 1991; Lai-Yin & Rob, 2005).

Second, Engel (1970) and Shermer (2008) found that humans typically do not intuit probability correctly without formal training. In fact, researchers have found that humans' lack of intuition regarding probability poses one of the primary difficulties in both the learning and teaching of the subject (Engel, 1970; Kahneman et al., 1982; Kahneman & Tversky, 1972, 1973a, 1973b, 1983; Shaughnessy & Bergman, 1993). This stumbling block creates a significant barrier to understanding abstract mathematical concepts. For example, when students examine the conjunction of two sets ($A \cap B$), they often conclude that the conjunction has a greater magnitude than the parent sets (Agnoli, 1987; Agnoli & Krantz, 1989; Shaughnessy, 1992). This intuitive response has been traced to the use of judgment heuristics such as representativeness (Agnoli, 1987; Shaughnessy, 1992). Heuristics are often useful for framing mathematical reasoning to solve problems (Pólya, 1957; Schoenfeld, 1992). However, reliance on heuristic judgments may lead to reasoning errors when they reflect beliefs rather than attributes or when those heuristics are used as substitutes for understanding the meaning of concepts (Kahneman & Tversky, 1983; Shaughnessy, 1992). Consider an example in which students are asked to compare two sets, "Men who have had one or more heart attacks" and "Men who are over 55 years old" (Agnoli, 1987, p. 3). Basing the comparison on representativeness beliefs often leads students to conclude that the conjunction of these two sets, "Men who have had one or more heart attacks and are over 55 years old," is actually larger than either of the two parent sets (i.e., they believe that men over 55 are more likely to have a heart attack) when, actually, parent sets are always larger than their conjunction.

Third, the misconceptions prevalent in probability may influence the foundations of how students think about mathematics generally. For example, Green (1983b) reported on a survey of over 3,000 British teens' (ages 11 – 16) in which he identified a major misconception in the area of proportions: Students most commonly chose the incorrect answer that corresponded with the largest numerator value rather than the relative size of the rational number relationship. Through tasks such as Green's Marble Problem, an exploration of probability problems provides a natural venue for exploring rational number concepts. The authentic experiences so necessary for learning probability concepts also require scrutiny of the meaning of the rational numbers used for reporting probabilities. Additionally, the concrete applications of probability (e.g., flipping coins, rolling a die, simulation and modeling) may be a critical key for students struggling to integrate abstract concepts into their prior conceptions of mathematics (Evans & Tsatsaroni, 2000; Freudenthal, 1970; Fuson, 1998; Green, 1983b; Watson & Shaughnessy, 2004). Connecting abstract concepts such as randomness to probability simulations such as *The Cliff Hanger* applet (Mathematics, Science, and Technology Education, 2005) may also improve student orientations toward mathematics (Stone, Alfeld, & Pearson, 2008) and student flexibility in unfamiliar problem solving situations (Evans & Tsatsaroni, 2000). As a result, students may develop a deeper relational understanding of mathematical concepts, allowing them to handle greater mathematical complexity and difficulty.

A Conceptual Framework to Model Mathematics Learning

The similarity between reasoning across mathematics content areas suggests a pathway of learning that either results in understanding or misconceptions and errors. To develop a model that traces these pathways, several factors must be considered. First, the introduction of new concepts is typically accompanied with tasks or problems for the students to complete. The characteristics of these tasks (e.g., task complexity, difficulty, discrimination between ability levels) may influence how students interpret the new material. Second, students must filter tasks through their own knowledge framework. Third, the pedagogical emphasis on either concepts or procedures direct students to develop either relational or instrumental understanding (Skemp, 1976/2006). If students learn relationally, then the conceptual understanding they develop may produce stronger, more consistent procedural skills, which in turn may reinforce deeper more robust conceptual understanding. This understanding may then be integrated into a student's knowledge framework for use with future tasks.

Alternatively, the development of instrumental understanding leads to the development of procedures without meaning, with incomplete or erroneous meaning, or even the lack of awareness of meaning (Skemp 1976/2006). Misunderstanding the meaning of mathematical objects in some way is the very essence of misconceptions. Mathematical misconceptions result in errors that are often difficult for teachers to prevent or obstruct. Researchers have repeatedly found that systematic errors due to misconceptions rather than faulty reasoning adhere to patterns of over- or under-generalization of properties or concepts for a particular task (e.g., Chang, 2002; Falk, 1992; Fuys & Liebov, 1997; Kalchman & Koedinger, 2005; Van Dooren, De Bock, Depaepe, Janssens, and Verschaffel, 2003). For example, in geometry, Fuys and Liebov (1997) suggested that students struggling to move from a visualization level of spatial reasoning to an analysis level may under-generalize geometric properties by including irrelevant characteristics of a shape in their mental framework; or conversely, they may over-generalize relationships between figures by discarding any number of a shape's unique properties. If unchecked, these misconceptions may be integrated into students' mathematical understanding, thereby influencing future learning.

Difficulties Inherent to Addressing Mathematical Misconceptions Directly

Multiple attempts to develop interventions for reducing misconceptions have met with limited success. Some of these efforts have focused on addressing task-specific errors (e.g., Rosnick & Clement, 1979). One difficulty with such a strategy is that if the error was due to a misconception, the underlying misconception will remain in the student's knowledge framework to adapt and reappear in the same or other task. Other endeavors have attempted to address the reasoning that leads to an error using a variety of strategies such as worked examples (e.g., Fisher, 1988; Phillippe, 1992; Rosnick & Clement, 1980). Directly addressing erroneous reasoning appeared to make no significant improvement in student learning (Weinberg, 2007).

Weinberg (2007) suggested that another reason student errors can be so insidious is that students attempt to adapt their knowledge base to the problem scenario, sometimes accurately and sometimes not. The adaptive nature of these errors suggests that the reasoning processes are built on a deeper foundation of understanding relating to the structure and meaning of mathematical ideas (Kieran, 2007, 2008, 2009).

An even more robust intervention design may be needed to alter students' mathematical thinking and reasoning. Rather than targeting reasoning processes directly, such an intervention might focus instead on transforming the instrumental understanding responsible for difficulties in meaning that can lead to misconceptions into relational understanding. If a teaching intervention targets the development of meaning and connections, then misconceptions that develop may be only a normal, temporary part of the learning process (Resnick, 1983).

Mathematical Task Characteristics

Teachers typically introduce new concepts by presenting a task or problem as a motivation for learning the mathematical concept. Rousseau (1976) identified eight task characteristics that potentially influence how students internalize the meaning of the task and its connection to the underlying concept: task identity, task autonomy, skill variety, task variety, task feedback, task learning, dealing with others, and task significance. Task identity refers to the ownership a student assumes for an activity. Task autonomy, closely aligned with identity, focuses on the degree of independence students have in decision making throughout a task. Skill variety emphasizes the breadth and depth of skills required to complete a particular task. Task variety, on the other hand, refers to the breadth of subjects and courses provided by a school. Task learning represents the breadth and scope of opportunities for obtaining new skills, what Hiebert and Grouws (2007) referred to as opportunity to learn. Task feedback speaks to the amount of feedback students receive from a task versus the feedback from teachers. Catanzaro (1997) maintained that task feedback creates a more stimulating, positive learning environment over instructor feedback. Rousseau (1976) defined dealing with others as "the opportunity to interact with teachers, teaching assistants and other faculty" (p. 3).

Many researchers would also emphasize the importance of interactions with other students (e.g., Berg, 1993; Freeman, 1997; Henderson & Landesman, 1995; Nichols & Miller, 1994; Parham, 1993; Slavin & Karweit, 1982; Slavin & Lake, 2008; Slavin, Lake, & Groff, 2009; Whicker, Bol, & Nunnery, 1997). Task significance represents student perceptions of a particular task's relevance to life beyond academic concerns. Rousseau (1976) found that task significance may have the strongest impact of her eight task characteristics.

Student Thought Processes Influencing Mathematical Misconceptions

Erroneous thinking resulting from misconceptions is often stable and robust, interfering with a student's ability to learn mathematics (Moschkovich, 1998). Researchers tend to agree that a possible key to addressing these issues may lie in the alignment of student thought processes with mathematical logic (e.g., Behr, 1980; Blanco & Garrote, 2007; Collis, 1975; Enfedaque, 1990; Kieran, 1980; Palarea Medina, 1999; Socas Robayna, 1997) and the connection of specific misconceptions to the student's larger knowledge framework (e.g., Moschkovich, 1998; Smith, diSessa, & Roschelle, 1993). This knowledge framework includes (at least) four components that can influence whether a student develops relational or instrumental understanding: (1) Discernment; (2) Orientation toward mathematics; (3) Individual context; and (4) Environmental Context.

Discernment. Discernment has been defined as an aspect of knowledge that encompasses the active, cognitive components of learning (Ronau & Rakes, 2010; Ronau, Rakes, Wagener, & Dougherty, 2009; Ronau, Wagener, & Rakes, 2009). Kant (1786/1901) proposed that cognition is engaged through the process of perceptions leading to conceptions, which in turn lead to ideas. Davis (1992), comparing Japanese to

American tests, considered the influence of such perceptions to be paramount to deep mathematical learning:

> Perhaps 75 one-step problems on a test will produce about the same ranking of students as will 6 multistep problems that require serious thought (and perhaps some originality). *But the message that they send to students is entirely different*. The one-step problems say to students, “You do not have to do much hard thinking in mathematics, nor must you be very creative; all you have to do is pay attention in class, memorize dutifully, practice diligently, and you will get no surprises on the tests.” The Japanese tests send a different message — rather more in the spirit of the contest problems that a very few U.S. students encounter — where it is more clear from the outset that, if you have developed nothing more than routine skills, you will be hopelessly ineffective. You *must* strive for ingenuity and originality (p. 725).

Davis (1992) went on to consider the meaning of mathematics from a cognitive perspective. He gave three examples of problems whose solution required the addition of whole numbers. These problems differed in the degree of decision making required about each contextual situation prior to concluding that addition is needed for each.

> Now, here is the main point behind these three examples: Most people who have not had an opportunity to think seriously about such matters would claim that the mathematics is that part of the problem that the calculator did. They might find the decisions…or the choice of arithmetical operations…to be thought provoking, but they would

> probably not consider them an essential part of the mathematics…they might not even notice that there was any thinking involved other than the computation that the calculator carried out. I would argue that such observers are precisely wrong. There is very little mathematics in the actual carrying out of the computations…The mathematics lies mainly in analyzing the real situation and deciding how to represent it in an appropriate abstract symbolic form (Davis, 1992, p. 727).

Schoenfeld (1992) agreed with Davis' conceptualization of the nature of mathematical learning. He added that mathematical problem solving requires a great deal of metacognitive regulation and that such behavior is learned best through "domain-specific instruction" (p. 357). In an earlier work (Schoenfeld, 1982), he considered three types of analysis to be important to mathematical problem solving: analysis of tactical knowledge (i.e., domain-specific facts and procedures), analysis of control knowledge (i.e., strategic/executive behavior), and analysis of belief systems. The analysis of control knowledge speaks directly to metacognition, the regulation of cognitive processes. Several other researchers have suggested that cognitive and meta-cognitive skills filter student ability to understand mathematical concepts (Andrade & Valtcheva, 2009; Dermitzaki, Leondari, & Goudas, 2009; Fuson et al., 2005; Lin, Schwartz, & Hatano, 2005; Nemirovsky & Ferrara, 2009; Usher, 2009). Swanson (1990) found that the development of metacognition may operate independently of aptitude and may impact learning more:

> On the surface, it appears that high metacognitive skills can compensate for overall ability by providing a certain knowledge about

> cognition. This knowledge allows low-aptitude/high-metacognitive children to perform in ways similar to those of children with high aptitude. Thus, one may argue that measures of metacognition and general aptitude in the present study are tapping different forms of knowledge, and that high performance on the problem-solving tasks is more closely related to higher performance on the metacognitive measures than on the aptitude measures (Swanson, 1990, p. 312).

Schraw and Dennison (1994) identified two constructs that measure metacognition: knowledge of cognition and regulation of cognition. They found that, although the two constructs are correlated, each may affect cognitive performance in a unique way. Other studies have shown that students use these cognitively-based discernment faculties to connect abstract concepts to concrete representations (e.g., Secada, 1992; Spillane, 2000; Von Minden, Walls, & Nardi, 1998).

Orientation toward mathematics. Schoenfeld's (1982) third type of analysis focused on student beliefs. He posited that student beliefs about the nature of a mathematical task can greatly influence the degree of cognitive effort expended for the task. Schoenfeld (1985) conducted a survey of 230 students in three high schools. He found three aspects to student beliefs about mathematics. (1) Students in his sample attributed success in mathematics to work rather than luck. (2) Students in his sample disagreed that mathematics solutions were either "right" or "wrong." They also declared the importance of teaching multiple ways to solve mathematics problems. This response surprised Schoenfeld because "very little of such teacher behavior was observed in the classroom studies…their response suggests either a strong acceptance of the mythology about teaching, or some strong degree of wishful thinking" (p. 14). (3) Students view mathematics learning as largely dependent on memorization while simultaneously viewing it as a means to develop logical thinking.

McLeod (1992) agreed with Schoenfeld's description of beliefs and attitudes as components of affect; however, he added a third, distinct component category: emotions. Emotional reactions to mathematics learning occur when students experience obstacles to solutions. Such obstacles elicit negative feelings such as tension, frustration, fear, anxiety, embarrassment, and panic. Once obstacles are overcome, positive emotions return. He maintained that one goal of mathematics pedagogy should be to reduce the occurrence of these negative emotions. From attitudes, beliefs, and attitudes, seven subconstructs of affect emerge: confidence, self concept, self efficacy, anxiety, effort and ability attributions, learned helplessness, and motivation.

Schoenfeld and Mcleod agreed that affect and cognition are linked (Schoenfeld, 1989; McLeod, 1992). Schoenfeld (1989) found that beliefs and attitudes influence the way people develop conceptions about mathematics, directly and indirectly impacting their mathematical ability. Barkatsas, Kasimatis, and Gialamas (2009) found that high levels of mathematics achievement are associated with positive attitudes toward learning mathematics; positive attitudes, in turn, are associated with mathematics confidence and affective engagement. Ismail (2009) found that self-confidence appeared to supersede the impact of socio-economic disadvantage on student achievement.

In summary, components of affect such as beliefs, attitudes, and emotions mold student orientations toward mathematics. Pedagogical strategies within mathematics influence the development of conceptions or misconceptions as a result of their attention to orientation.

Individual context. The contextual factors that students bring to a mathematical learning situation interact in multiple ways to influence how students interpret mathematical concepts. These individual context factors refer to characteristics such as gender, race, culture, socio-economic status, parent education levels, background experiences, and learning styles (Ronau et al., 2009; Ronau & Rakes, 2010; Ronau, Wagener, & Rakes, 2009).

Evidence has suggested that boys and girls construct their understanding of mathematics differently (Fennema & Sherman, 1977) and hold different attitudes toward mathematics (Sherman & Fennema, 1978). Although moderate changes have occurred over time, inequity between genders still exists (Carrell, Page, & West, 2009; Fennema, 2000; Mendick, 2008; Van Langen, Rekers-Mombarg, & Dekkers, 2008; Wei & Hendrix, 2009; Zohar & Gershikov, 2008).

Kozol (1992, 2005) examined educational practices across the country and asserted that inequalities also continue to exist across racial lines. Snipes and Waters (2005) agreed with Kozol’s assessment, conducting a case study in a single state. Lubienski (2001) and Lim (2008) found that race and class interact to produce an effect on mathematics achievement. Class measures include factors such as parent education levels and socio-economic status (SES). Parent education levels, one measure of SES, significantly predicted above average achievement during the Third International Math and Science Study (TIMSS; Schreiber, 2000). Lehrer, Strom, and Confrey (2002) found that prior mathematical experiences influence student orientation toward mathematics. Anderson (1990) asserted that cultural influences overshadow gender and racial effects on equity in student achievement. Nelson, Joseph, & Williams (1993) agreed with Anderson, claiming that culture also has a direct bearing on affect. Strutchens (1995) proposed the use of a five-dimensional framework for increasing equity in mathematics education: content integration, knowledge construction, prejudice reduction, equitable pedagogy, and empowering school and social culture.

Alomar (2007) and Esposito Lamy (2003) linked gender, race, culture, and affect with family variables such as parenting style and poverty. Lopez, Gallimore, Garnier, and Reese (2007) found that for immigrant populations, family factors influence English language literacy, which in turn affects student mathematics achievement.

Personal characteristics such as learning styles, personality, and temperament also influence how students learn mathematics. The Silver-Strong studies (Silver, Brunsting, & Walsh, 2008; Silver, Strong, & Perini, 1997; Strong, Perini, Silver, & Thomas, 2004; Strong, Silver, & Perini, 2001) together with the work of Keirsey (1998) suggest a link between learning styles and personality. Keirsey (1998) described personality in terms of the Myers-Briggs notation. In this framework, a person may be Introverted (I) or extraverted (E); rely more on intuition (N) or the senses (S) to interpret a situation; rely more on feelings (F) or thinking (T) to make decisions; and, prefer routine (J for judgment) or sponteneity (P for perceiving), resulting in 16 different personality styles that he grouped into four categories with internal reliability ratings between 0.82 and 0.83 (Alpine Media Corporation, 2003). Silver et al. (1997) used the same constructs to determine four categories of learning styles: Mastery, Understanding, Interpersonal, and Self-Expressive Learners. The dependence on these two frameworks on the Myers-Briggs constructs (Myers, 1962) suggests a possible link between learning styles and personality.

These frameworks directly map onto one another (Table 7).

Table 7
Alignment of Keirsey Personality Framework with Silver-Strong Learning Styles

Silver-Strong Learning Style	Values and Educational Preferences	Associated Keirsey Personality Types
Mastery	Value: Clarity and Practicality Prefer: procedure, drill and practice, concrete, closed questioning.	Guardian Administrators: ISTJ; ESTJ Artisan Operators: ISTP; ESTP
Understanding	Value: Logic and Evidence Prefer: logic, debate, inquiry, indendent study, argumentation, and why questions.	All Rational Subgroups: INTJ; INTP ENTJ; ENTP
Interpersonal	Value: The ability to help others Prefer: topics that affect lives, cooperative/collaborative learning, and teacher attention to successes and struggles.	Guardian Conservators: ISFJ; ESFJ Artisan Entertainers: ISFP; ESFP
Self Expressive	Value: Originality and aesthetics Prefer: use of imagination to explore ideas, creative artistic activity, open-ended questions, and generating possibilities and alternatives.	All Idealist Subgroups: INFJ; INFP ENFJ; ENFP

Understanding the role of values and preferences of the various types of learners directly impacts the equitable teaching of mathematics (Gardner & Hatch, 1989). Second, the traditional mathematics education described by Fey (1979) that still continues today (Hiebert, 2003; Hiebert & Grouws, 2007; Stigler & Hiebert, 1997) targets mastery learners almost exclusively, while they account for only about 35% of the population (Silver et al., 1997). In smaller samples, such as a single high school, the mastery learners have been found to account for far lower percentages (24% in Tungate, 2008). That mathematics teachers tend to be mastery learners themselves seems likely and would account for the disproportionate bent toward traditional practices.

Personality has been framed most prominently as five major constructs known as “The Big Five:” Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience (Ahadi & Rothbart, 1994, p. 189) . Personality emerges from temperament, but assessment of adult personality may not map directly from temperament (Rothbart, Ahadi, & Evans, 2000). For example, cognitive self-concept may supersede temperamental tendencies (i.e., beliefs about how a person would like to be, should be, and is in reality are difficult to separate).

Posner and Rothbart (2007), Rothbart and Jones (1998), Rueda, Rothbart, Saccomanno, and Posner (2007) and Rudasill (2009) asserted that Attention, one temperament factor, may influence the learning of mathematics both directly and indirectly. “Everywhere in cognitive neuroscience, specific brain networks seem to underlie performance. However, some of those networks have the improtant property of being able to modify the activity in other networks” (Posner & Rothbart, 2007, pp. 15-16).

In brief, individual factors such as gender, race, class, personality, learning styles, and background experiences interact to influence orientation and cognition in mathematics. Moreover, evidence suggests that temperament may be a critical individual learning factor. Equitable mathematics teaching requires the consideration of the unique effects of these individual context factors.

Environmental Context. Environmental factors interact with individual factors to influence the equitability of learning opportunities in mathematics. Controversy over the importance of environmental factors on learning lasted for decades, beginning with the publication of *Equality of Educational Opportunities*, more commonly known as *The Coleman Report* (Coleman et al., 1966). This study examined the achievement impact of differences between races on: school factors such as class size, access to chemistry, physics, and language laboratories, number of books in libraries, number of textbooks; teacher and principal characteristics such as type of college attended, years of teaching experience, salary, maternal education level, vocabulary ability, and dispositions; and student characteristics such as parental background, presence of parents at home, size of family, parental expectations, parental involvement, and socio-economic status. *Equality* fundamentally altered definitions of equality from simply comparing resource "inputs" to analyzing the effects of inputs on educational achievements (Coleman, 1967a). Coleman (1967b) considered the complexity of implicit assumptions present within an input-based notion of equality:

> It is one thing to take as given that approximately 60% of an entering high school freshman class will not attend college; but to assign a particular child to a curriculum designed for that 60% closes off for that child the opportunity to attend college. Yet to assign all children to a curriculum designed for the 40% who will attend college creates inequality for those who, at the end of high school, fall among the 60% who do not attend college… there is a wide variety of different paths that adolescents take on the completion of secondary school (Coleman, 1967b, p. 9).

Instead, *Equality* examined inequality based on five different criteria: degree of racial segregation, allocation of resources, teacher orientations, weighted resource inputs based on achievement predictability, and output (e.g., achievement, career choice) differences (Coleman, 1968). *Equality* found that student characteristics accounted for the majority of variance in achievement and of the impact of teacher characteristics on learning. For example, Coleman et al. (1966) reported that teacher variables accounted for 2.06% of the variation in mathematics achievement for Black students but only 0.61% for White students (p. 294). They concluded that "variations in school quality are not highly related to variations in achievement of pupils" (p. 297). However, technology capabilities of the time limited the researchers' analytic capabilities (Stringfield & Teddlie, 2004). Later studies (e.g., Bryk & Raudenbush, 1988; Raudenbush & Bryk, 1984) took advantage of technological advancements by conducting multilevel analyses on the subsets of the Coleman et al. (1966) data set:

> The results were startling — 83% of the variance in [learning] growth rates was between schools. In contrast, only about 14% of the variance in initial status was between schools…this analysis identified substantial differences among schools that conventional models would not have detected (Raudenbush & Bryk, 2004, pp. 9-10).

Recent studies have continued to emphasize the importance of environmental factors on students learning. Hegedus and Kaput (2004) found that the way classroom activities are organized affects the potential depth of student understanding. Cobb, Gresalfi, and Hodge (2009) found that cultures within a classroom influence the development of personal identities in mathematics. LaRocque (2008) found that student perceptions of the classroom environment are associated with reading and mathematics achievement. She noted that the interaction of perception with gender was not statistically significant but that the interaction of perception with grade level was significant. McMahon, Wernsman, and Rose (2009) agreed with LaRocque's findings that perceptions of classroom difficulty are strong predictors of mathematics and science self-efficacy. Bong (2008) found that classroom goal structures influence student perceptions of mathematics learning. She also found that relationships influence perceptions of learning. In like manner, Carter (2008) described the impact of having a classroom climate that values the struggle of connecting mathematical concepts to current conceptions. She concluded that such a climate enhances student self efficacy and confidence. Similarly, Murayama and Elliot (2009) concluded that classroom goal structures influence the development of intrinsic motivation.

Amenkhienan and Kogan (2004) concluded that the student-teacher relationship influences the amount of learning that occurs. Stemler, Elliott, Grigorenko, & Sternberg (2006) proposed a framework for interpersonal relationships with teachers, noting that the work of teaching is largely social in nature. Likewise, Hughes and Kwok (2007) identified teacher relationships with both parents and students as mediating factors of student motivation and achievement. They also noted an interaction between race and the amount of teacher support received. Osterman (2000) summarized research findings on the interaction of student belongingness and school and classroom conditions with motivation and achievement:

> Research also tells us that conditions in the classroom and school influence students' feelings about themselves; these in turn are reflected in student engagement and achievement. Not all students experience alienation to the same extent, yet, for the most part, students and researchers describe schools as alienating institutions…While the "peer culture" may establish norms dress and behavior, it is not necessarily one that satisfies students' need for belongingness (p. 360).

Stipek (2006) added to Osterman's findings:

> Learning requires effort, and one of the best predictors of students' effort and engagement in school is the relationships they have with their

> teachers…To promote high academic standards, teachers need to create supportive social contexts and develop positive relationships with students (p. 46).

Accordingly, the impact of the learning environment and student perceptions of that environment interact with individual context but also act as a distinct component to student learning.

Putting the Model Together

Figure 10 offers a pictorial interpretation of how the characteristics of a task and of a student's knowledge framework may operate within a mathematics classroom learning environment. Procedural knowledge isolated from conceptual knowledge and the connections between ideas results in instrumental understanding (Skemp, 1976/2006). Instrumental understanding may result in a cycle of misconceptions and faulty reasoning reinforcing each other and weakening a student's knowledge framework for understanding future tasks. When conceptual knowledge and procedural knowledge develop together, they reinforce each other and strengthen a student's knowledge framework for future tasks. When students complete a task, teachers have limited opportunities to assess the knowledge framework and thought processes that lead to a response; instead, assessment usually focuses on whether or not a response was correct. Unfortunately, correct responses do not necessarily indicate that a student understands the mathematical concepts completely. Figure 10 therefore includes the possibility that correct responses can be produced even with erroneous reasoning, and if unchecked, that reasoning will reinforce misconceptions and erroneous reasoning, thereby weakening a student's knowledge framework for understanding future tasks.

Resnick (1983) suggested that errors often occur when students look for meaning in situations where the given information is incomplete. In such cases, students often attempt to use their prior knowledge to fill in the gaps and yielding misconceptions. Thus Resnick (1983) found that working through these difficulties may be a normal part of the learning process and that combating misconceptions and faulty reasoning must become an expected part of the struggle that is so critical to deep conceptual learning as Hiebert and Grouws (2007) later pointed out. Moschkovich (1998) agreed with Resnick when she noted refinement of understanding as a primary goal of teaching: "We need to understand the process of conceptual change that enables learners to transform and refine their conceptions to more closely fit with the desired understanding" (p. 209).

Tracing the root causes of errors and recognizing erroneous reasoning requires an examination of student explanations about their reasoning processes. Previous interventions targeting specific errors or the underlying reasoning have met with limited success (e.g., Clement, 1982; Fisher, 1988; Phillippe, 1992; Rosnick & Clement, 1979), possibly because these interventions may have targeted the error instead of the latent reasoning and misconception that led to the error. Furthermore, students with misconceptions and faulty reasoning may produce correct answers; as a result, interventions focusing on errors may miss unobservable erroneous reasoning.

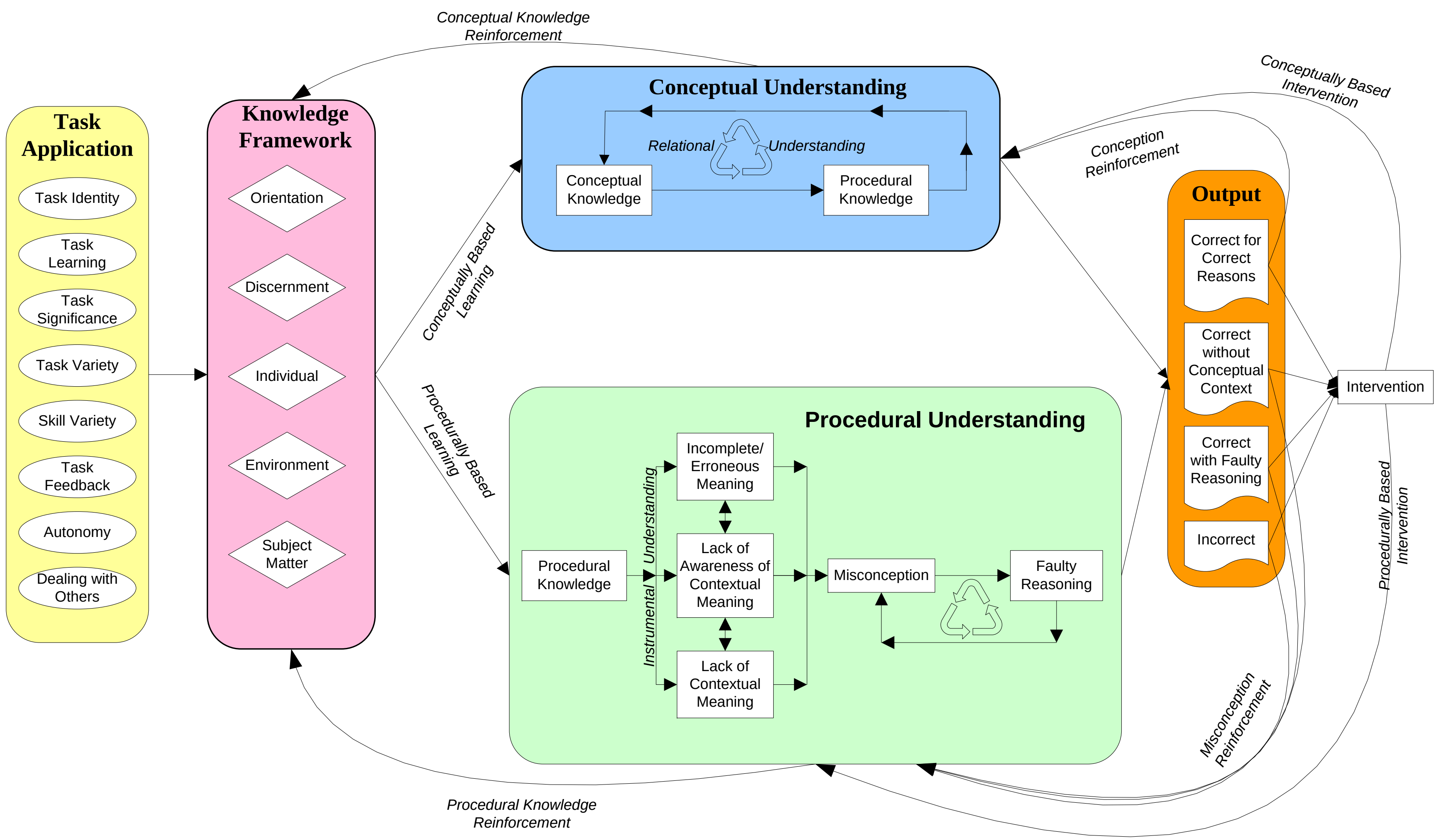


Figure 10. Pathways of Mathematical Learning.

The rationale for using probability instruction as an intervention targets these concerns: (1) Focusing on connections between probability concepts and algebra, geometry, and rational number concepts may help students develop relational understanding; (2) Probability instruction focuses on developing meaning rather than eliminating errors; therefore, if students produce correct answers for incorrect reasons, the development of meaning may help alter the misunderstandings that led to the errorneous reasoning; and, (3) Probability simulations and experiments offer concrete explorations for students investigating complex, abstract mathematical phenomena that often lead to misunderstanding, misconception, and faulty reasoning.

Summary and Research Questions

Research into misconceptions is necessarily problematic due to the latent nature of those misconceptions. Researchers must rely on observable errors and discern whether those errors are due to faulty reasoning despite having solid relational understanding of concepts or if they are due to misunderstanding about the meaning and connections of mathematical ideas.

Students encounter special difficulties when transitioning from whole numbers to rational numbers and from arithmetic to algebra. They often struggle to determine when linear proportions are or are not appropriate. They also have difficulty distinguishing between additive and multiplicative relationships. These difficulties appear in the study of rational numbers, algebra, geometry, and probability and often create obstacles to constructing meaning and connecting the meaning of ideas within structured relationships. While these difficulties are equally poignant in the learning of probability, the concepts within probability offer significantly more opportunities for simulations and experiments that help bridge the gap between abstract ideas and concrete examples.

Probability, however, is often deleted or minimized from the curriculum due to time constraints and/or teacher insecurity with the material. The interconnectedness of common mathematical misconceptions across probability, algebra, and geometry coupled with the limited training students receive in probability and the significant potential of probability experiments to bridge abstract mathematical concepts with concrete examples suggests the possibility that probability instruction holds the key to alleviating fundamental mathematics misconceptions. To explore this potential, the present study will examine the following four research questions:

1) Do probability misconceptions have a causal influence on algebra, geometry, and rational number misconceptions?
2) Does probability instruction reduce critical misconceptions in probability, rational numbers, algebra, or geometry?
3) Do student attitudes toward mathematics influence the emergence of errors due to misconceptions on mathematical tasks?
4) Does student metacognition influence the emergence of errors due to misconceptions on mathematical tasks?

CHAPTER 3

METHODOLOGY

The present study examines the structure of mathematical misconceptions in high school and the impact of attitudes toward mathematics and metacognitive knowledge and skills on the development of misconceptions. Additionally, the possible role of probability instruction as an intervention for mathematical misconceptions will be assessed.

The measurement of mathematical misconceptions is inherently problematic due to the latent nature of those misconceptions. For example, Zawojewski and Shaughnessy (2000) pointed out the inadequacy of simple multiple choice tests to identify the thought patterns that result in a particular answer. Instead, they recommended including a qualitative component to each question to provide clues to underlying student thinking. In order to include that strategy in the instruments used in this study, an initial assessment of student responses was necessary to determine the source of student reasoning errors. For example, were reasoning errors occurring on a particular due to a lack of relational understanding, despite having relational understanding, or due to a more fundamental misunderstanding of foundational mathematical ideas? The results of that analysis, presented in Chapter 4, were used to code errors for the subsequent quantitative analyses. As such, the design of this study falls within the mixed methodology design as described by Tashakkori and Teddlie (1998).

This chapter contains a description of the probability unit that will serve as the intervention along with the design of the study and its rationale, threats to validity, the assessment instruments and their reliability coefficients, and data analysis techniques.

Research Design

Subjects

The present study was conducted with 19 mathematics teachers recruited from four schools in three Kentucky school districts with 1,142 students enrolled in their 53 algebra and geometry classes. All elements of the protocol were approved by the University of Louisville's Internal Review Board for the protection of human subjects in research, as required by federal regulations (Protection of Human Subjects, 2009).

Design Description

The present study used a randomly assigned untreated control group with a pretest and switching replication (Equation 14; Shadish et al., 2002).

$$\frac{\begin{matrix} R & O_{ATMI/MAI} & O_{NAEP} & X & O_{ATMI/MAI} & O_{NAEP} & & O_{ATMI/MAI} & O_{NAEP} \end{matrix}}{\begin{matrix} R & O_{ATMI/MAI} & O_{NAEP} & & O_{ATMI/MAI} & O_{NAEP} & X & O_{ATMI/MAI} & O_{NAEP} \end{matrix}} \qquad (14)$$

The outcome of interest for the analysis of the intervention is the rate of growth during the treatment period; the collection of pretest data removes pre-existing differences as a source of group difference. Use of a control group allowed a comparison of growth rates in the intervention group and the normal rate of growth without the intervention thereby minimizing history and maturation threats to validity. Classes for each teacher were randomly assigned to treatment conditions using Microsoft Excel 2007 and Minitab 15 statistical software to minimize any selection threats to validity. The switching replication fulfilled two purposes: (1) The ethical obligations of research demand that all students receive the intervention instruction; (2) The post-post test provided data on the retention of intervention effects for follow-up studies.

Probability Instruction Intervention

The intervention for this study was a probability unit designed to provide students multiple opportunities to explore the meaning of fundamental mathematical concepts rather than targeting specific error patterns. Because the relationship between probability and algebra differs from that of probability and geometry, the probability concepts studied in each class varied.

In algebra, the intervention consisted of five lessons. The length of instruction varied across schools and teachers to accommodate the dynamics of particular classes. From a teaching perspective, such adjustments should be encouraged since students learn at varying paces. On the other hand, a research design perspective recognizes that such adjustments pose a history threat to internal validity. For example, events occurring concurrently with the treatment could cause the observed effect rather than the intervention itself (Shadish, Cook, & Campbell, 2002). Since the alternative of rigorously abiding by a timeline would also have increased the risk of multiple threats to validity, I chose to stay in close communication with each teacher about adjustments made to the timeline. For most teachers, the intervention lasted approximately ten 90-minute class periods. The overall topics for these lessons were:

1. Statistical structure (Appendix A and B)
2. Randomness (Appendix C and D)
3. Counting principles (Appendix E and F)
4. Event probability (Appendix G, H, and I)
5. Probability distributions (Appendix J and K)

In geometry, the intervention consisted of three lessons, normally lasting approximately six 90-minute class periods. The overall topics for these lessons were:

1. Counting principles (Appendix E and F)
2. Geometry probability (Appendix G and I)
3. Probability distributions (Appendix J and L)

Teachers were also provided lesson plans for classes in the control condition; most, however, chose to continue with their normal instructional sequence. Boston & Smith (2009) suggested that teacher-made materials may not offer students the same degree of cognitive load. Furthermore, teacher-made materials increase the potential unreliability of treatment implementation threat to statistical conclusion validity (Shadish et al., 2002). To manage this threat, classes were randomly assigned within teachers so that teacher effects were distributed across both to treatment conditions. Classroom observations and teacher interviews were conducted to measure the degree of heterogeneity between groups.

Instrumentation

Three instruments were used to measure student mathematics knowledge, student attitudes toward mathematics, and student metacognitive knowledge and skills. The mathematics knowledge instrument was used to account for pre-existing mathematics knowledge and ability. It was also used to analyze error response patterns to determine which errors emerged from mathematical misconceptions or from non-conceptual reasoning errors.

Mathematics Knowledge Instrument

Items for the mathematics knowledge instrument (Appendix N) were gathered from National Assessment of Educational Progress (NAEP) released items (U.S.Department of Education, 1996, 2005, 2007). Although all 17 items remained as given by NAEP, a prompt was included for each question asking students to explain how or why they chose their response.

NAEP items are rigorously developed, using review boards, pilot testing, classical test theory, and Item Response Theory to analyze item performance (U.S.Department of Education, 2008a). These items were deemed to have high content validity for the NAEP-associated content areas.

These items included rational number, probability, algebra, and geometry content. Table 8 provides a description of each NAEP item used in the assessment instrument along with the reported reliability coefficients for each item block (U.S. Department of Education, 2008b, 2008c, 2008d, 2008e). These items were chosen based on two criteria: (1) The item content matched the foundational concepts that research has suggested connect rational number, probability, algebra, and geometry misconceptions closely enough to be able to detect intervention effects; and, (2) The item content and wording did not so closely match the activities and problems in the probability unit that the treatment group would receive an unfair advantage over the control group. Table 8 presents the classical test theory difficulty coefficient (i.e., percent correct), the NAEP classification of difficulty and complexity level of each item with respect to the intended grade level of the item, and the internal consistency of the associated block of items as they appeared on the NAEP instruments.

.

Table 8

Reported NAEP Item Performance

Item	Release Year	Content Strand	Content	Percent Correct	Grade Level	Difficulty	Complexity	NAEP Block	Cronbach Coefficient α
1	2007	Probability	Relative versus absolute comparison	45%	Grade 4	Medium	Low	M7	0.80
2	2005	Probability	Determine Conditional Probability	49.5%	Grade 12	Medium	Low	M12	0.73
3	2007	Probability	Repeated Sampling Probability	60%	Grade 8	Medium	Low	M11	0.76
4	2005	Probability	Dependent probability	18%	Grade 8	Hard	Moderate	M12	0.75
5	2007	Algebra	Convert temperature units	35%	Grade 8	Hard	Low	M9	0.80
6	2005	Algebra	Effect of variable change	34%	Grade 8	Hard	Moderate	M3	0.76
7	1996	Algebra	Additive versus Multiplicative Structure	58%	Grade 8	Medium	-	M3	0.53
8	2007	Algebra	Solve algebraic word problem	47%	Grade 8	Medium	Moderate	M11	0.76
9	2007	Geometry	Determine if a shape is a parallelogram	26%	Grade 8	Hard	Moderate	M11	0.76
10	2005	Geometry	Area of shaded figure	77%	Grade 8	Easy	Low	M4	0.77
11	2005	Geometry	Find dimensions from scale drawing	85%	Grade 12	Easy	Moderate	M12	0.75
12	2005	Rational Number	Rational Number Quantity Meaning	66%	Grade 12	Easy	Low	M3	0.73
13	2005	Rational Number	Given the scale, determine length of side	56%	Grade 12	Medium	Low	M4	0.79
14	2007	Rational Number	Arrange fractions in ascending order	49%	Grade 8	Medium	Low	M9	0.80
15	2007	Rational Number	Determine fraction of figure shaded	89%	Grade 8	Easy	Low	M11	0.76
16	2007	Algebra	Determine equation to represent table.	54%	Grade 8	Medium	Moderate	M7	0.78
17	2005	Probability	Determine amount from probability	40%	Grade 8	Medium	Low	M4	0.77

Since the items were chosen from different blocks, the NAEP-reported coefficients do not necessarily represent the internal consistency of the new instrument compiled for the present study. Therefore, the pooled internal consistency of the new instrument was re-assessed using the pretest data (α = 0.791, 95% CI [0.773, 0.808]) and the posttest data (α = 0.772, 95% CI [0.751, 0.773]) and found to have adequate internal consistency. The correlation of each item (Table 9) between the pre- and post-tests were computed to measure test-retest reliability (i.e., stability). The correlations were moderate and significant ($p < 0.001$) for all items except Item 17, which was only significant at the 93% confidence level ($p = 0.068$). Overall, the stability of the items appeared to be acceptable (Table 9).

Table 9
Stability Correlations Between the Pre- and Post-Test Data for each Item

Item	1	2	3	4	5	6	7	8	9	10	11
Correlation	0.491	0.173	0.273	0.310	0.277	0.217	0.422	0.308	0.279	0.354	0.385
Item	10	11	12	13	14	15	16	17			
Correlation	0.354	0.385	0.268	0.154	0.460	0.300	0.325	0.083			

Content validity of content area alignment to national, state, and local standards was evaluated by the NAEP Validity Studies Panel. Daro, Stancavage, Ortega, DeStefano, & Linn (2007) examined the content coverage, skill coverage, alignment to NAEP framework, lack of philosophical bias, lack of ability bias, and representativeness of information provided about students. They found that 96% of NAEP 2005 and 2007 items demonstrated adequate or marginal quality.

Item Response Theory (IRT) was applied to measure the characteristics of difficulty, discrimination (i.e., the ability to distinguish between groups, in this case, ability levels), and guessing for each item. IRT, unlike Classical Test Theory (CTT), focuses on the correctness or incorrectness of each item individually rather than a raw cumulative score (Baker & Kim, 2004). In CTT, difficulty is defined as the percentage of correct responses for an item (as in Table 12, means for pretest items and Table 13, means for posttest items). CTT discrimination is typically measured as the point-biserial correlation for each item. One problem with CTT is the circular dependence of observed scores and samples (Fan, 1998).

IRT is based on the item characteristic curve, which is computed using a logistic function. The curve can be computed as a Rausch model (1 parameter, item difficulty), 2PL (2 parameters, item difficulty and discrimination), or 3PL (3 parameters, item difficulty, discrimination, and guessing). The logistic function for the 3-PL curve is

$$P(\theta) = c + (1 - c)\frac{1}{1 + e^{-a(\theta - b)}} \tag{15}$$

where:

a represents the discrimination coefficient

b represents the difficulty coefficient

c represents the guessing coefficient

θ represents the ability level of the respondent

ParScale 4.1 (Muraki & Bock, 2002) uses an iterative process to compute the item characteristic curve. In the first iteration, ability levels (θ) for each subject were computed. These values become the starting point for the second iteration, which is used to compute the values for *a*, *b*, and c. The guessing coefficient, *c*. was estimated as c = 0 for all 17 items. Therefore, the model reduced to a 2PL curve, and the values for a and b were computed for each item (Table 10).

Table 10
IRT Coefficients for NAEP Items

Item	Discrimination, *a*	Difficulty, *b*
1	0.570	-0.618
2	0.630	0.379
3	0.693	-0.205
4	0.962	-0.090
5	0.827	-0.175
6	0.570	0.930
7	1.192	-0.273
8	0.730	0.430
9	0.428	-1.308
10	0.918	-0.446
11	1.224	-0.080
12	0.879	0.149
13	0.876	0.246
14	0.861	0.311
15	0.883	-0.554
16	1.036	0.203
17	0.744	0.525

Note: SE for all 17 items for both a and b was < 0.001

The item characteristic curves (Figure 11) can be used to compare the behavior of each item. For each curve, the horizontal axis represents difficulty, *b*, and the vertical axis represents the ability, θ.

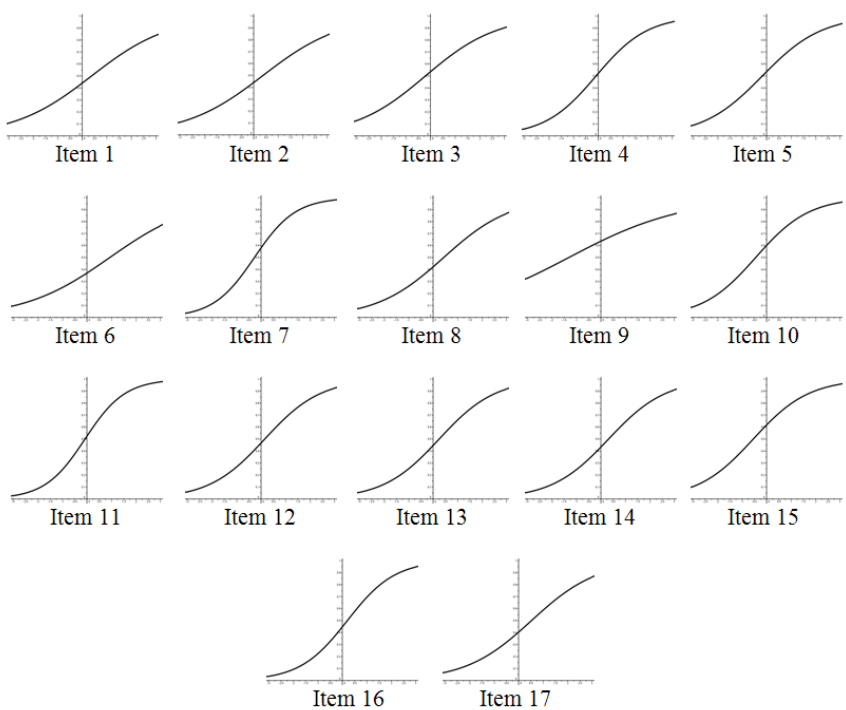


Figure 11. Item Characteristic Curves for NAEP Mathematics Knowledge Instrument

None of the curves in Figure 11 leveled off at the top or bottom, indicating no ceiling or floor effects. The difficulty of an item is defined as the point on the item characteristic curve for which the ability level is average, θ = 0.5. The discrimination of the item, *a*, is defined as the slope of the curve. Being a highly difficult item does not necessarily mean that an item is also highly discriminating across ability levels. Consider a comparison of Items 6 and 7 (Figure 12).

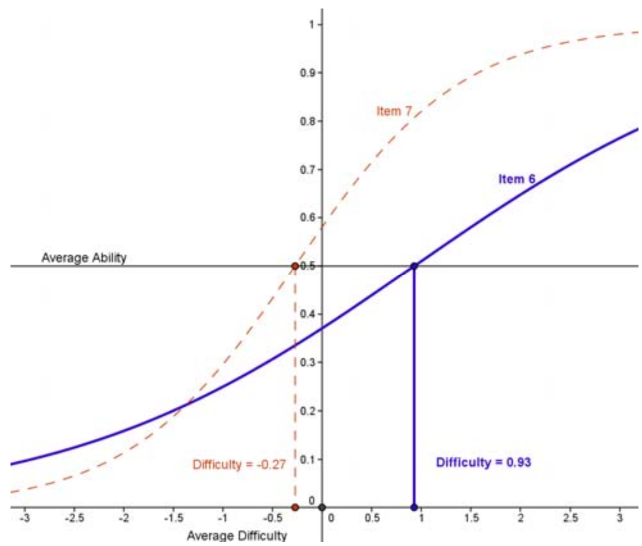


Figure 12. Comparison of Item Curves for Items 6 and 7

The difficulty of Item 6 was 0.93 while the difficulty of Item 7 was -0.27, so Item 6 was the more difficult of the two items. The slope, however, of Item 7 (1.192) was steeper than the slope of Item 6 (0.570); this characteristic means that although Item 6 was more difficult than Item 7, Item 7 discriminated between ability levels more than Item 6.

The discrimination of items on the mathematics knowledge test ranged from 0.428 (less differences between high and low ability students) to 1.192 (more differences between high and low ability students). The difficulty ranged from -1.308 (Easy) to 0.930 (Hard). Because the discrimination levels did not approach 0 (no differences between high and low ability students), and the difficulty levels did not indicate that any items were extremely easy (approaching -3 and +3), the item characteristics were considered appropriate for the planned analyses.

Mathematics Attitudes Inventory

Student orientation was measured using the Attitude Toward Mathematics Inventory (ATMI; Appendix O; Tapia & Marsh, 2004). This instrument was selected because its subscales have been extensively analyzed to establish high reliability and content validity (Tapia & Marsh, 2004). The subscales for this inventory were developed from multiple literature sources to maximize concurrent construct validity. According to the Tapia and Marsh report, the ATMI measures four orientation constructs. Factor 1, self confidence, consists of 15 items with a reported Cronbach alpha of 0.95. Factors 2 and 3, perceptions of the value of mathematics and enjoyment of mathematics, each contain 10 items with a Cronbach alpha of 0.89. Factor 4, motivation to learn mathematics, contains five items with a Cronbach alpha of 0.88.

The internal consistency for the full instrument and each subscale was measured using the present study data to determine their reliabilities (Table 11).

Table 11
Internal Consistency Reliability for ATMI

Scale	Observed Cronbach Alpha	95% Confidence Interval
Full Instrument	0.943	[0.938, 0.949]
Factor 1: Self Confidence	0.909	[0.901, 0.918]
Factor 2: Value	0.876	[0.864, 0.888]
Factor 3: Enjoyment	0.798	[0.778, 0.817]
Factor 4: Motivation	0.824	[0.806, 0.842]

The observed reliability coefficients for the present study data appeared to be comparable to those reported by Tapia and Marsh (2004) and had values higher than the typical threshold of 0.7 (Urbina, 2004). The reliabilities were, therefore, determined to be acceptable.

MetaCognition Inventory

Student metacognition knowledge and skills were measured using the Metacognitive Awareness Inventory (MAI; Appendix P; Schraw & Dennison, 1994). This instrument was selected because of its unique subscales of metacognition, knowledge of cognition and regulation of cognition and because it has been rigorously tested through two experiments to establish concurrent construct validity for each block of items. Three types of knowledge are measured as components of knowledge of cognition: (1) Declarative knowledge, defined as knowledge of learning and of one's own cognitive skills and abilities; (2) Procedural knowledge, knowledge of how to use various cognitive strategies; and, (3) Conditional knowledge, knowledge of when to use particular cognitive strategies and why those strategies should be used. Under the regulation of cognition, five components are measured: (1) Planning, including goal setting and allocation of resources; (2) Organizing and managing information; (3) Monitoring, reflection on cognitive processes during a learning task; (4) Debugging, strategies for correcting performance errors or assumptions; and, (5) Evaluation, reflection on cognitive processes after a learning task is completed (G. Schraw, personal communication, May 31, 2009). Schraw and Dennison (1994) also reported high internal consistency for the whole instrument ($\alpha = 0.93$) and both metacognition factors ($\alpha = 0.88$). The internal consistency for the full instrument and each subscale was measured using the present study data to determine their reliabilities (Table 12).

Table 12

Internal Consistency Reliability for MAI

Scale	Observed Cronbach Alpha	95% Confidence Interval
Full Instrument	0.946	[0.941, 0.951]
Factor 1: Knowledge of Cognition	0.870	[0.857, 0.882]
Declarative Knowledge	0.744	[0.718, 0.768]
Procedural Knowledge	0.615	[0.573, 0.654]
Conditional Knowledge	0.668	[0.633, 0.700]
Factor 2: Regulation of Cognition	0.924	[0.917, 0.932]
Planning	0.727	[0.699, 0.752]
Organizing	0.788	[0.767, 0.808]
Monitoring	0.735	[0.708, 0.760]
Debugging	0.694	[0.662, 0.725]
Evaluation	0.673	[0.639, 0.705]

The observed reliability coefficients for the full instrument and two main factors demonstrated high internal consistency. Several of the sub-factors showed marginal reliabilities ($\alpha < 0.7$), so only the two main factors were used in the subsequent analysis of contextual factors.

Missing Data

Rubin (1987) classified missing data due to non-response as either unit non-response, meaning that the subject refused to answer any of the items, and item non-response, meaning that the subject skipped questions.

Unit and Item Non-Response

The ATMI and MAI surveys of the present study included both types of non-response. Forty six students did not respond to any items on either survey; 63 additional students did not respond to a majority of items, ending at various points throughout the survey (Table 13). The format of the questionnaire may shed light on the most typical pattern of unit non-response: The front of the survey form included the ATMI and the first three questions of the MAI. Questions 4 – 52 of the MAI (i.e., the back of the survey) were the most commonly skipped questions. Based on this pattern, which may well be the result of bias in the non-response patterns, I concluded that most non-response on the MAI was due to the presentation format of the instrument.

The NAEP achievement data, both pre- and post-test, consisted of very low proportions of missing data. On the pretest, missing data accounted for 5.6% of the entries across all items and subjects, and only 11 students (< 1%) did not respond to any items (Table 13). On the posttest, missing data accounted for 6.5% of the entries across all items and subjects, and only 12 students (1.1%) did not respond to any items.

Table 13
Sources of Missing Data

Instrument	N	Unit Non-Response (No items answered)	Item Non-Response
ATMI	964	46 (4.6%)	• 0 of 40 items with full data • 256 (26.6%) cases missing at least one value • 2,457 of 16,388 (6.4%) values missing
MAI	964	109 (11.3%)	• 0 of 52 items with full data • 316 (32.8%) cases missing at least one value • 6,060 of 16,388 (12.1%) values missing
Pretest	1142	11 (0.96%)	• 0 of 17 items with full data • 242 (21.4%) cases missing at least one value • 1093 of 19,278 (5.6%) values missing
Posttest	1021	12 (1.1%)	• 0 of 17 items with full data • 248 (24.3%) cases missing at least one value • 1121 of 17,357 (6.5%) values missing

Imputation of Missing Data

Multiple imputation is an expansion of multiple regression imputation that uses Bayesian inference from observed data using probability models to impute values for the missing data. The process of multiple imputation begins by estimating a probability model for the observed data (the prior distribution). The process continues by computing a conditional probability distribution based on the observed data (Gelman, Carlin, Stern, & Rubin, 2004). First, the model regresses each missing data point on every other variable. Second, the true value for the missing data point is considered the mean of a distribution. Sampling error will therefore result in a potentially different value each time a multiple regression imputation is run. To account for this variance, multiple imputation creates any number of complete data sets. Rubin (1987) suggested that three to ten imputation sets are needed to account for variance in missing data. Brick, Jones, Kalton, and Valliant (2005) and Garson (2009) suggested that five sets are typically used. Each data set is used in subsequent analyses, and the results of each analysis are averaged. To complete the overall analysis, standard errors for each resulting point estimate are computed. Table 14 presents the sample sizes, means, and standard deviations for each imputed data set for the pretest and surveys, which were administered at the same time. *Imputation 0* represents the unimputed data set after listwise deletion.

Table 14
Means and Standard Error for Pretest and Survey Data Set

Imputation	5	4	3	2	1	0
Pretest Misconception Responses						
Sample Size (N)	1133	1133	1133	1133	1133	900
Mean Ratio	0.361	0.361	0.363	0.362	0.362	0.352
Standard Error	0.006	0.006	0.0063	0.006	0.006	0.007
T ratio from Imputation 0	1.381	1.381	1.652	1.534	1.534	-
Pretest Correct Responses						
Sample Size (N)	1133	1133	1133	1133	1133	900
Mean	0.480	0.481	0.481	0.481	0.480	0.495
Standard Error	0.007	0.007	0.007	0.007	0.007	0.007
T ratio from Imputation 0	-2.143*	-2.000*	-2.000*	-2.000*	-2.143*	-
ATMI Enjoyment						
Sample Size (N)	921	921	921	921	921	918
Mean	2.960	2.959	2.950	2.959	2.962	2.959
Standard Error	0.026	0.026	0.026	0.026	0.026	0.026
T ratio from Imputation 0	0.038	0.000	-0.346	0.000	0.115	-
ATMI Motivation						
Sample Size (N)	921	921	921	921	921	911
Mean	2.953	2.952	2.947	2.952	2.956	2.957
Standard Error	0.030	0.030	0.030	0.030	0.030	0.031
T ratio from Imputation 0	-0.131	-0.164	-0.328	-0.164	-0.033	-
ATMI Value						
Sample Size (N)	921	921	921	921	921	919
Mean	3.528	3.527	3.527	3.529	3.530	3.532
Standard Error	0.025	0.025	0.025	0.025	0.025	0.025
T ratio from Imputation 0	-0.160	-0.200	-0.200	-0.120	-0.080	-
ATMI Self Confidence						
Sample Size (N)	921	921	921	921	921	916
Mean	3.217	3.220	3.215	3.218	3.219	3.219
Standard Error	0.026	0.026	0.026	0.026	0.026	0.026
T ratio from Imputation 0	-0.077	0.038	-0.154	-0.038	0.000	-
MAI Knowledge of Cognition						
Sample Size (N)	921	921	921	921	921	893
Mean	3.423	3.427	3.422	3.435	3.424	3.443
Standard Error	0.021	0.021	0.021	0.021	0.021	0.022
T ratio from Imputation 0	-0.930	-0.744	-0.976	-0.372	-0.883	-
MAI Regulation of Cognition						
Sample Size (N)	921	921	921	921	921	894
Mean	3.214	3.223	3.216	3.220	3.213	3.221
Standard Error	0.019	0.019	0.020	0.019	0.019	0.020
T ratio from Imputation 0	-0.359	0.103	-0.250	-0.051	-0.410	-

*$t > 1.96$, $p < 0.05$

In addition to the means and standard deviations displayed in Tables 12, Figures 13 – 20 display the frequency distribution for each data set for each variable.

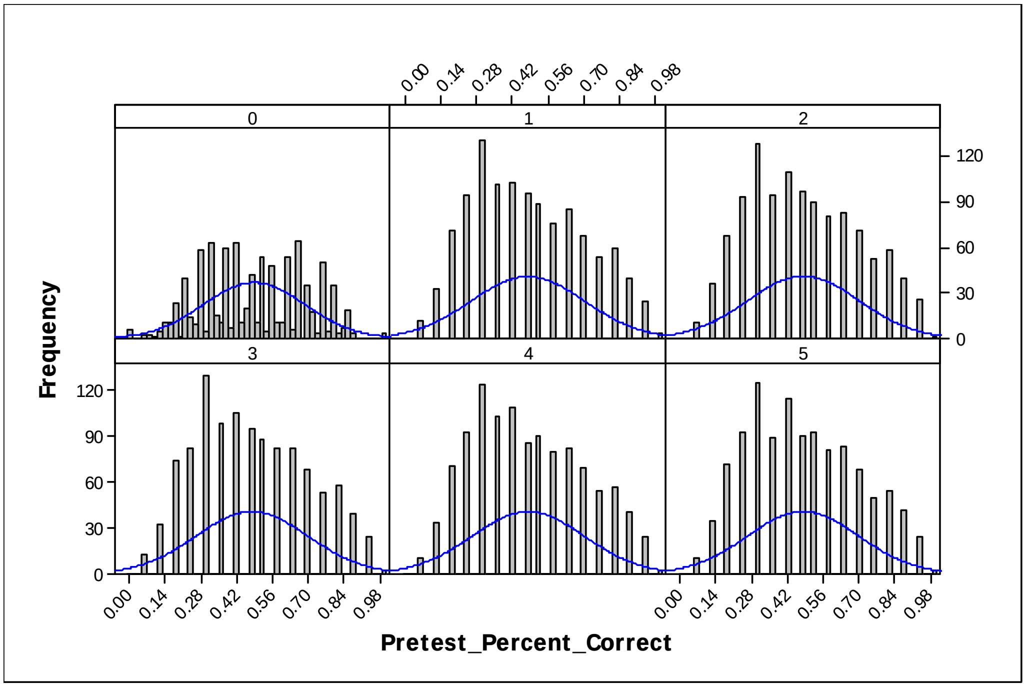


Figure 13. Pretest Percent Correct Data Distributions.

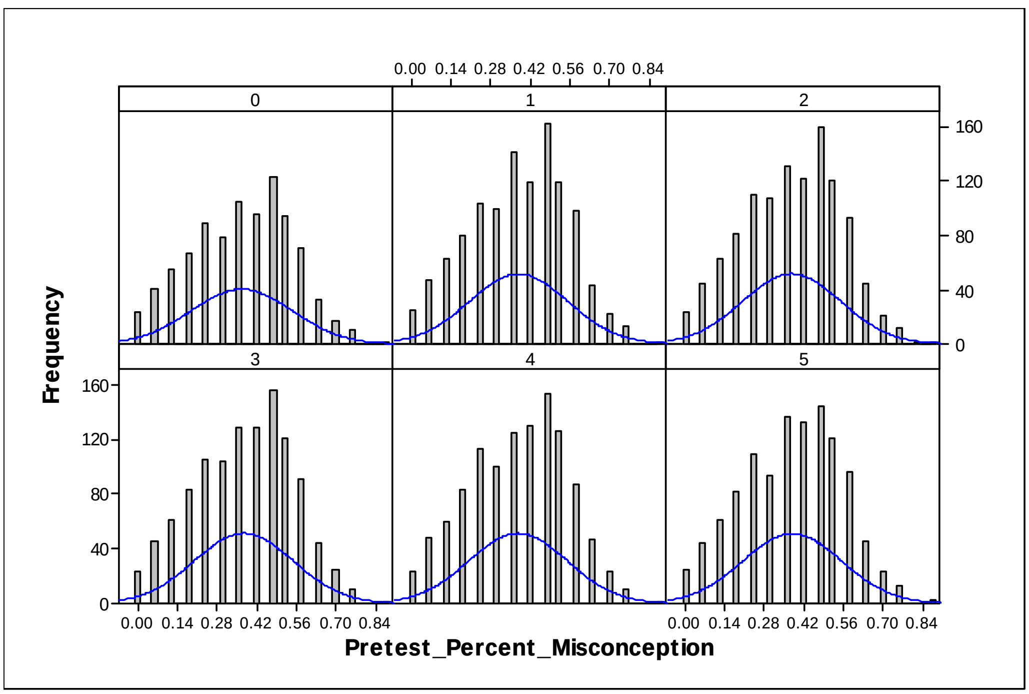


Figure 14. Pretest Percent Misconception Data Distributions.

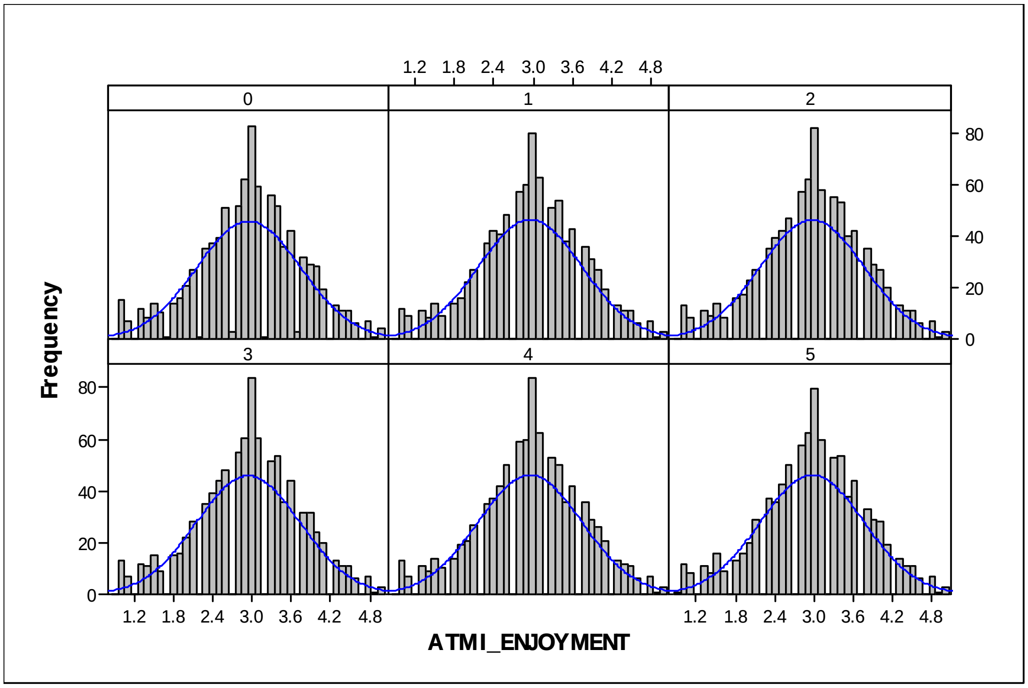


Figure 15. ATMI Enjoyment of Mathematics Data Distributions.

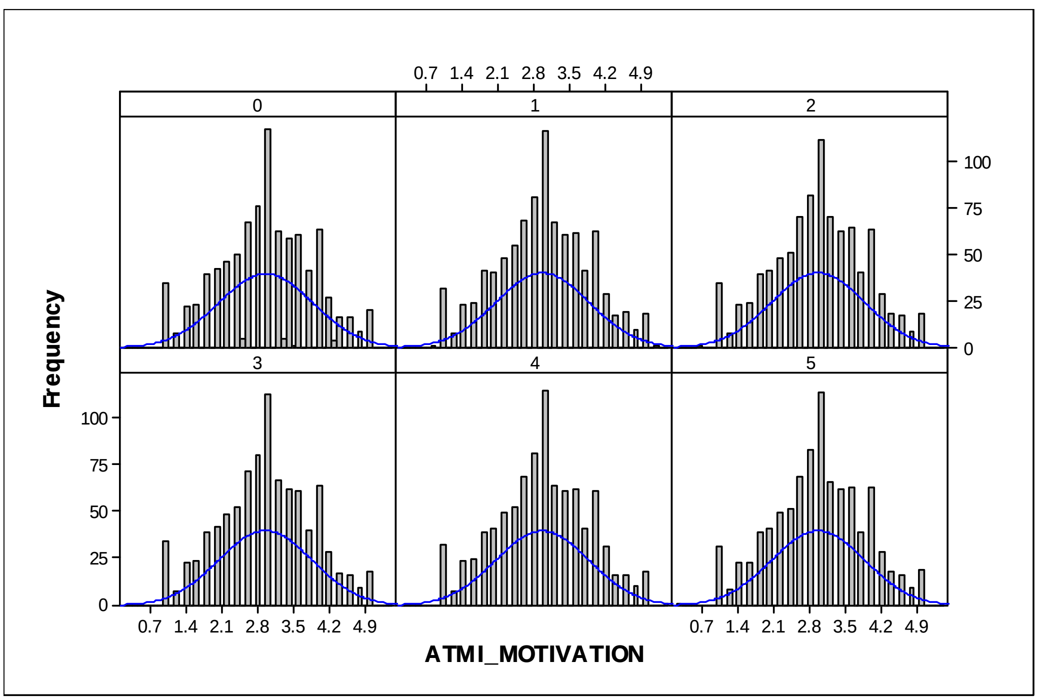


Figure 16. ATMI Mathematics Motivation Data Distributions.

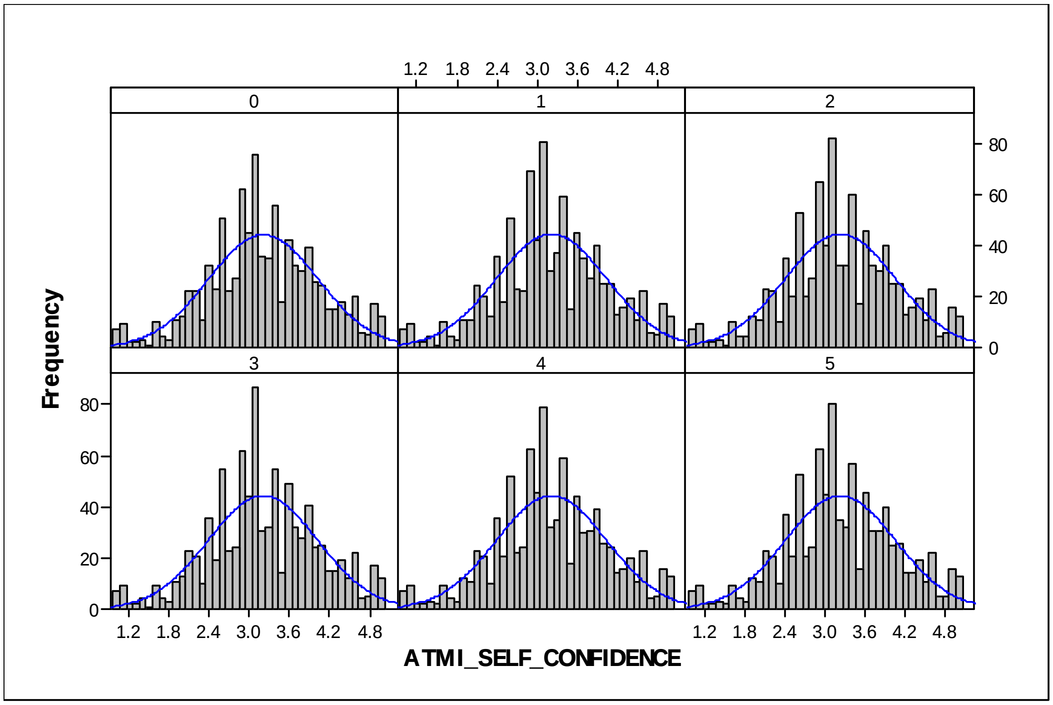


Figure 17. ATMI Mathematics Self Confidence Data Distributions.

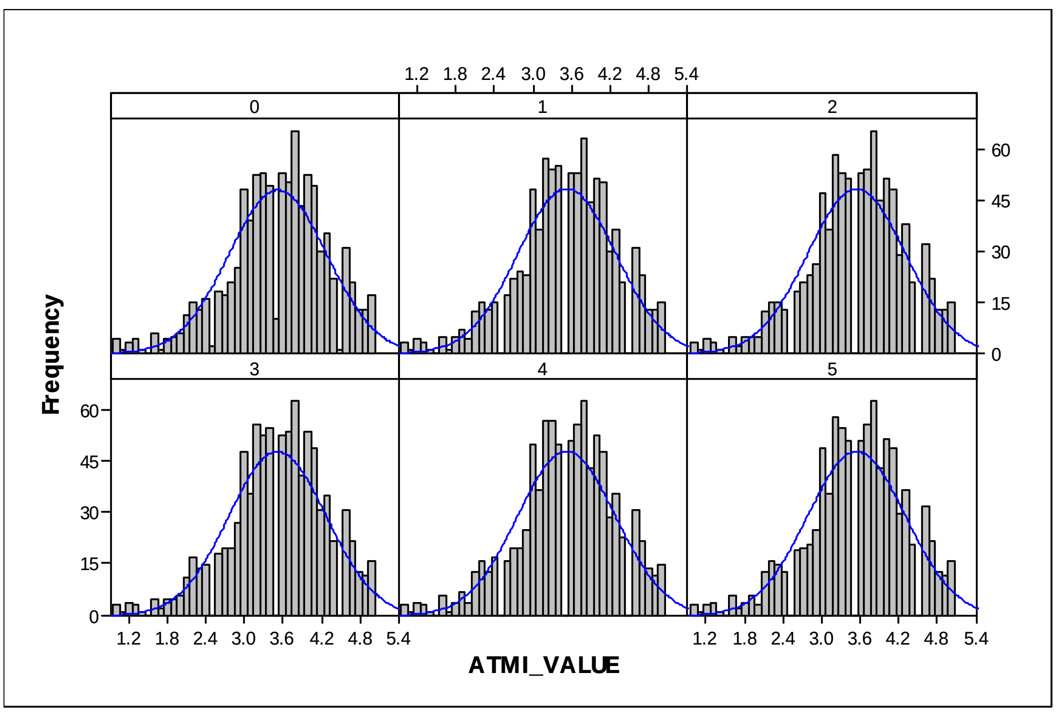


Figure 18. ATMI Value of Mathematics Data Distributions.

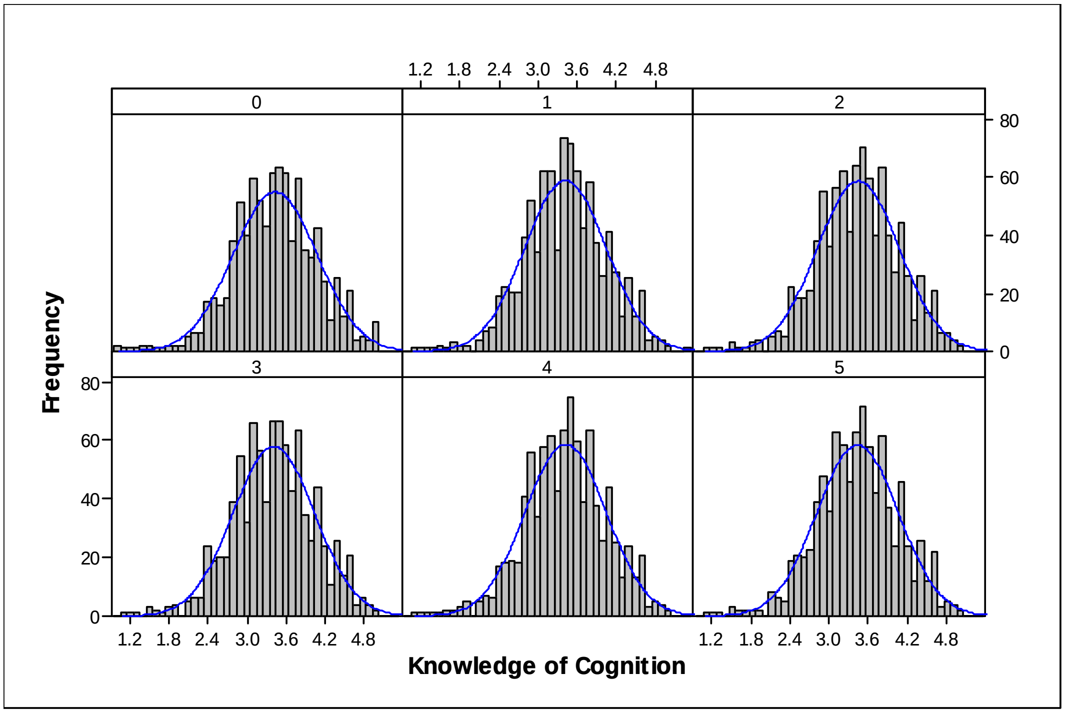


Figure 19. MAI Knowledge of Cognition Data Distributions.

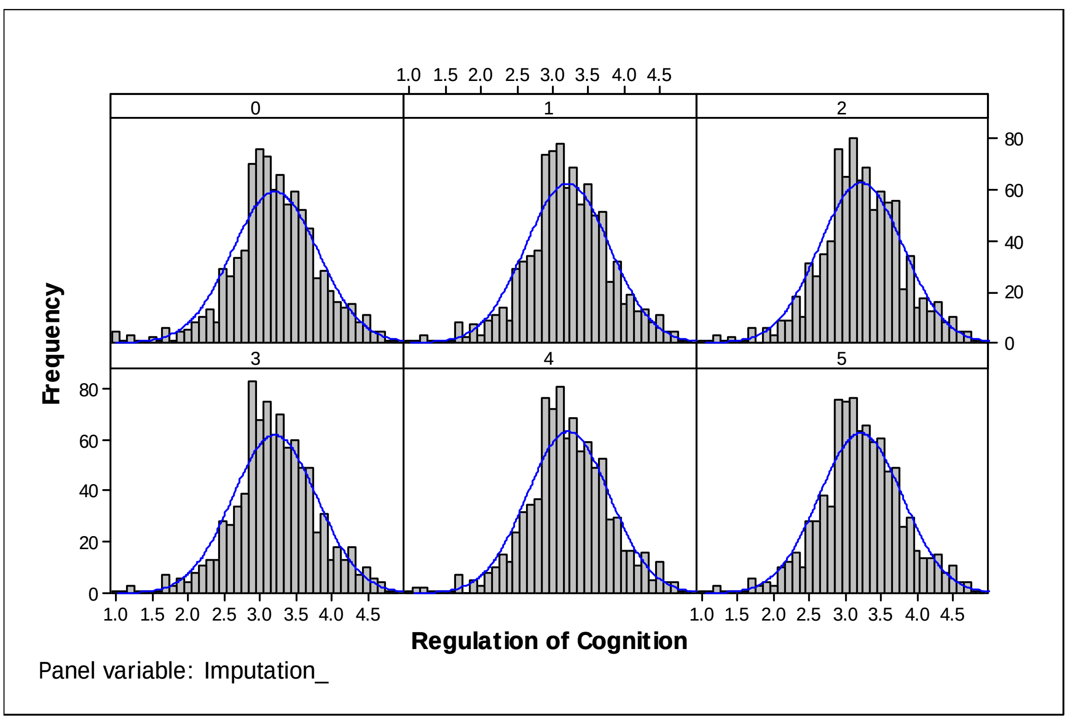


Figure 20. MAI Regulation of Cognition Data Distributions.

Missing data on the posttest were also imputed using multiple imputation. Table 15 displays the means and standard errors for these data sets, and Figures 21 and 22 present the frequency distribution for each posttest variable.

Table 15
Means and Standard Error for Posttest Data Set

Imputation	5	4	3	2	1	0
Misconceptions Responses						
Sample Size (N)	915	915	915	915	915	690
Mean Ratio	0.377	0.375	0.376	0.375	0.377	0.368
Standard Error	0.003	0.003	0.003	0.003	0.003	0.004
T ratio from Imputation 0	2.546*	1.980*	2.263*	1.980*	2.546*	-
Correct Responses						
Sample Size (N)	915	915	915	915	915	690
Mean	0.486	0.485	0.484	0.485	0.486	0.498
Standard Error	0.007	0.007	0.007	0.007	0.007	0.009
T ratio from Imputation 0	-1.488	-1.612	-1.736	-1.612	-1.488	-

*$t > 1.96$, $p < 0.05$

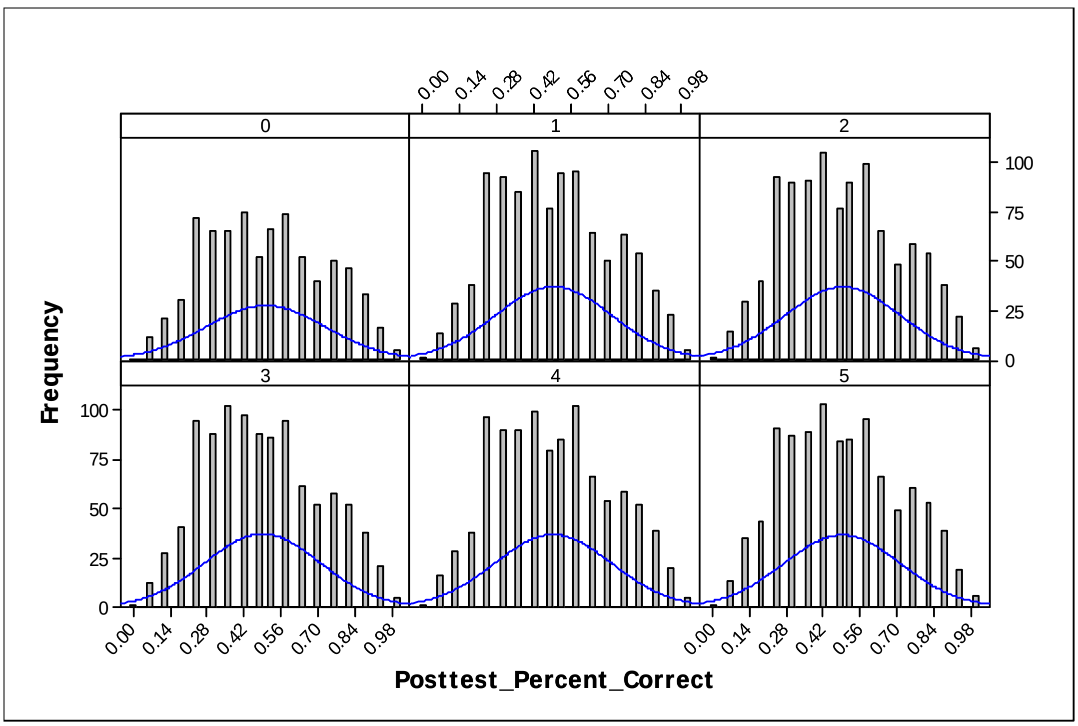


Figure 21. Posttest Percent Correct Data Distributions.

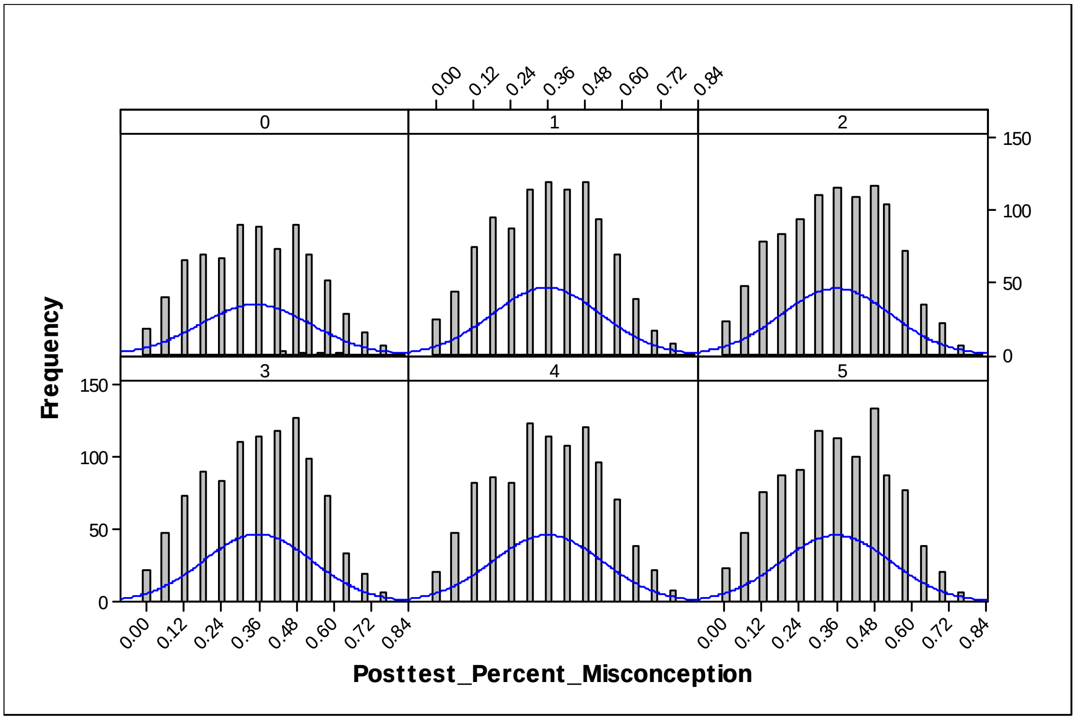


Figure 22. Posttest Percent Misconception Data Distributions.

T tests were used to compare each imputation to the raw data file for the pretest, surveys, and posttest. These tests indicated that the imputed data sets for the pretest and posttest variables contained significant differences from the unimputed data set (i.e., $|t| > 1.96$), implying that the data deleted because of unit and item non-response through listwise deletion contained important information that was lost. The ATMI and MAI factors did not appear to have significant differences between the unimputed and imputed data sets. Based on these results, the imputed data sets were used for all quantitative analyses, and the results were compiled into a mean and standard deviation for each statistic.

Attrition

In addition to unit and item non-response, the posttest consisted of additional missing data due to attrition. These data were un-recoverable through imputation because not enough data were available to make inferences about the posterior distributions of responses. Attrition occurred in three schools for different reasons, resulting in 113 students who took the pretest but not the posttest and approximately 100 students who took neither test.

In school A, one teacher refused to participate in the study without informing me, the department chair, or the principal. When I observed his classes, the instruction in both treatment and control classes matched the expected condition. Prior to instruction, the teacher informed me that he "needed to organize" the pretests. On subsequent visits, he was absent. As other teachers began to complete the posttest, I sent a message to the teacher asking when we could meet to hand off his data. I received the following message in response:

> I don't have any data for you. I never did your study because I didn't have a need to. I had already covered my stats that was required for Algebra 1 earlier in the year, and to repeat it would have put me way behind schedule for the semester. As for my Geometry classes, I work probability into each unit, and to talk about it as a separate unit did not seem reasonable for me. I had originally thought I would make up data for you, but then I realized two things, one that that isn't fair to you, and two, it was going to be too much work to make it up

Because these concerns were not voiced until the study was almost completed, addressing them in time to avoid the loss of his classes was impossible.

In school B, two teachers chose to leave the study because of pressure from the administration to increase their pace of instruction because of concerns about state testing. The three teachers from the first school and the two teachers from this school had already administered the pretest, so their data were retained for the qualitative analysis of student responses and the structural model analyses.

In School C, the superintendent of the district volunteered the entire mathematics department to participate in the study. Implementation of the protocol began with the administration of the pretest. Three teachers gave the pretest before the others. The day after these teachers had completed the pretest, the principal of the school required the department chair to withdraw the school from the study. No complaints about the study or the protocol from the mathematics teachers were responsible for this decision; rather, the difficulties appeared to be the result of internal disagreements at the district level. The following message was sent from the school's mathematics department chair:

> I have some bad news. Our principal called a math dept. meeting this morning to inform us that we would not be participating in the research study. I am not really sure what happened and I didn't even know we were having the meeting until she came in and had it announced that we were meeting. She said after speaking to some people in the dept., she doesn't want us to spend 6 days giving the test and survey's because it would take away from instruction. One teacher had already given the pre-test and survey and it took 3 days which will bump us up to 9 days for all three. It is because we have such short periods. We had a big setback in math test scores last year and the complaints were that we did not have enough time to cover the content. She and the assistant superintendent made the decision to pull us from the study because we are already behind in our curriculum again this year. I am very sorry.

In response to this message, I discussed the situation at length over the phone with the department chair, and as a result, I went to the school to meet with the principal and discuss her concerns. Although the principal was expecting me and had agreed to meet, when I arrived the principal refused to meet with me. Based on this response and subsequent conversations with the department chair and superintendent, I decided to halt further efforts to persuade the principal to remain in the study.

Statistical Power

In a meta-analysis examining instructional interventions in algebra (Rakes, Valentine, & McGatha, 2010), significant effect sizes in algebra across multiple intervention strategies averaged around 0.33. Based on that review of literature, this effect size was deemed to be a reasonable target when computing statistical power (i.e., the ability to detect significant effects). Tables from Cohen (1988) were consulted to find the minimum sample sizes needed to obtain a power of 0.80 (as recommended by Cohen, 1988, p. 390) for finding an effect size of 0.33. The tables in Cohen (1988, p. 384) recommended a sample size of 45 for an effect size of 0.30 and 33 for an effect size of 0.35. Using linear interpolation, the computed sample size needed to meet the power requirements was 38.

Statistical power in a cluster randomized experiment is influenced more by the number of clusters than by the number of subjects per cluster (Spybrook, 2008). The target sample of 38 was therefore directed toward the number of classes in the study rather than the number of students. Computations using Optimal Design Software (Raudenbush, 2009) confirmed this estimate (Figure 23).

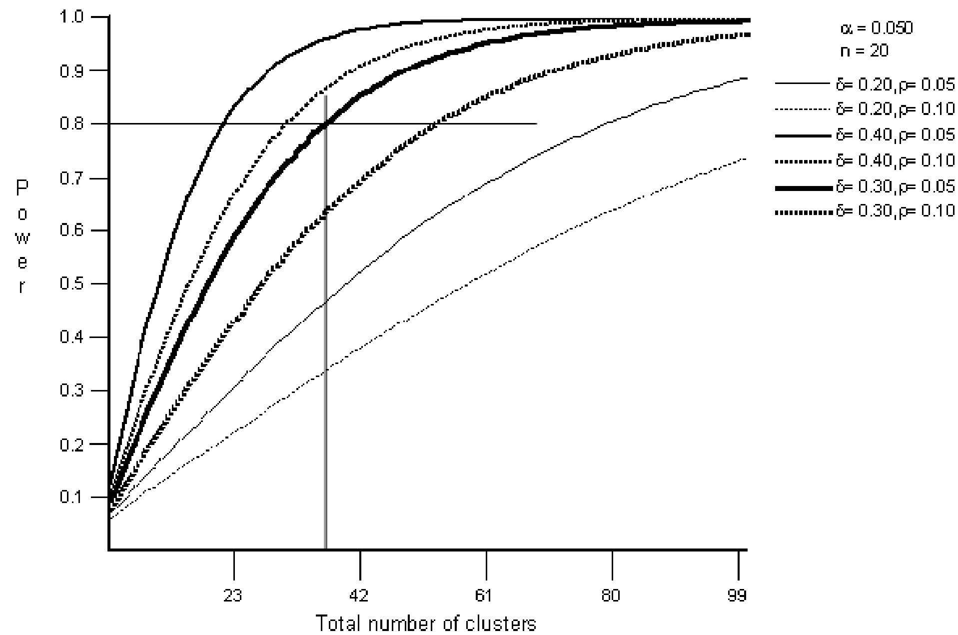


Figure 23. Power Curve for Cluster Randomized Trials

The curve in Figure 23 estimates the power based on an assumption of having an average of 20 students per cluster with an intra-class correlation of 0.05. These assumptions seemed reasonable: The average number of students in each class was 17 (SD = 4.3).

Data Analysis

The data analysis for the present study progressed through three stages: (1) qualitative analysis of student error responses and explanations, (2) structural analysis of content area misconceptions, and (3) hierarchical analysis of student and contextual factors on misconceptions. The qualitative analysis of student responses was used to adjust the coding of misconceptions for the subsequent quantitative analysis.

Qualitative Analysis

The qualitative analysis served two critical functions in the present study. First, classroom observations and teacher interviews (structured around topics relating to implementation of the intervention lessons) before, during, and after the treatment periods provided data on fidelity of treatment. While observing classrooms, the researcher attempted to minimize distractions inherent to having a visitor in a classroom. In some classrooms, this goal was best met by slipping in quietly and sitting in the back of the class. In other classrooms, teachers preferred to introduce the researcher and involve him in the lesson.

Fidelity to the probability intervention lessons varied widely between teachers. Shaughnessy and Bergman (1993) pointed out that many teachers are uncomfortable with probability content; varying responses to such discomfort were expected. Some teachers preferred to revert to normal, procedural methods of teaching. In this case, the ability of a probability unit to counter misconceptions may have been reduced. Other teachers followed the lessons provided by the researcher with varying degrees of success. One role of the researcher during the treatment period was to provide assistance to the teachers throughout the intervention lessons.

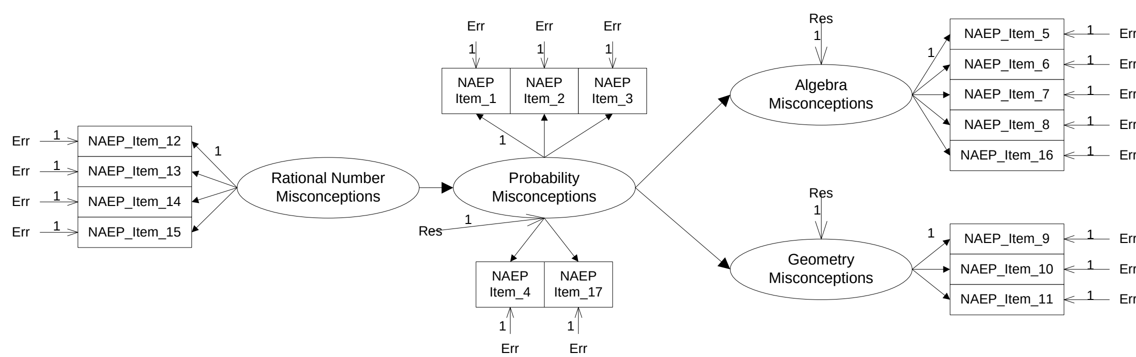
The second major function of the qualitative analysis was to provide an analysis of student responses to the open response items on the mathematics knowledge assessment. This analysis was used to assess hypotheses of previously identified misconception patterns and advance the understanding of the relationship between mathematical misconceptions and reasoning errors. This analysis was conducted from the constructivist point of view using a narrative analysis (Creswell, 2007; Patton, 2002) of symbolic interactions (i.e., the symbols used to provide meaning to students), semiotics (i.e., how signs and symbols are used to convey meaning), and hermeneutics (i.e., how students interpret signs and symbols).

These qualitative analyses were fundamental for establishing the context for all subsequent quantitative analysis. The first analysis provided evidence of treatment fidelity and intervention effectiveness. The second analysis provided a foundation for the interpretation of all quantitative findings.

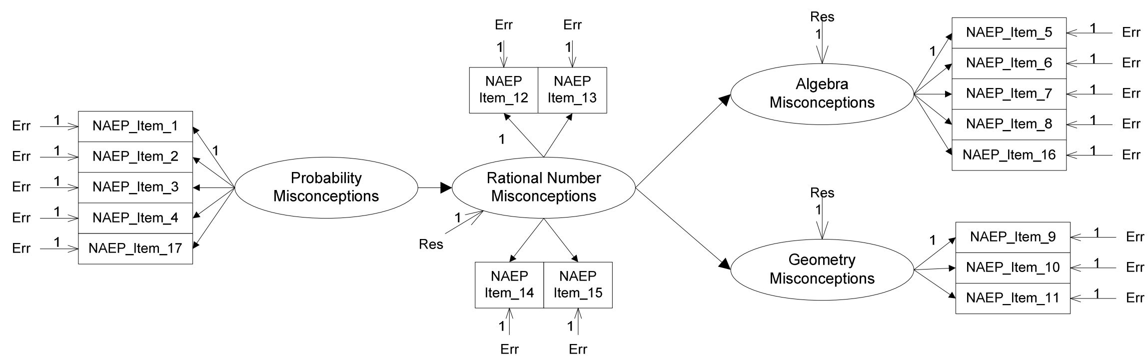
Quantitative Analysis

The quantitative analysis was carried out in two stages. First, using the pretest data on the NAEP multiple choice items, possible causal relationships between content area misconceptions were examined using structural equation modeling. Second, the impact of item, student, and class characteristics on misconception errors was investigated using hierarchical modeling.

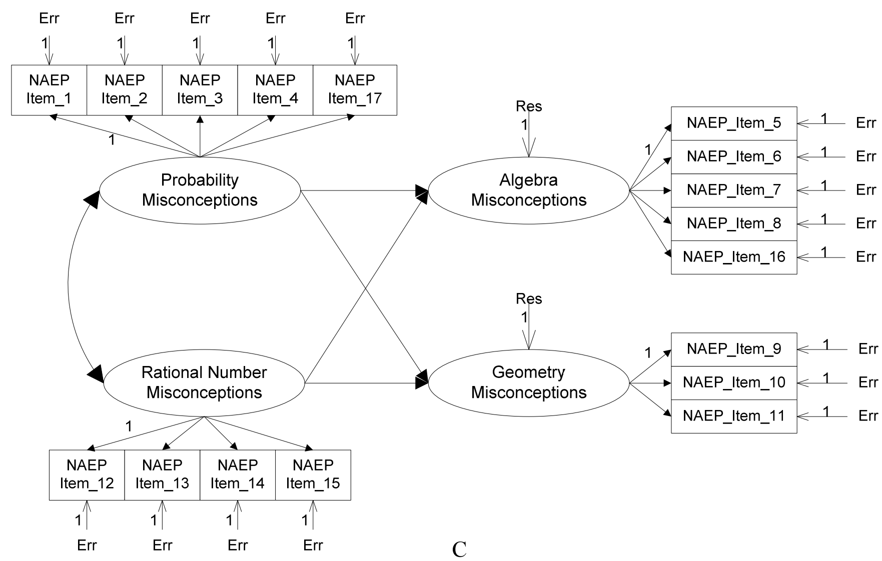
Structural analysis of content area misconceptions. Six structural equation models were used to compare the competing hypothesized models of misconception relationships among content areas (Figure 24). The pretest data were randomly split into two groups irrespective of treatment group assignment. The first group of pretest data was used to calibrate the six hypothesized models. The second group of pretest data was used to validate the resultant models.



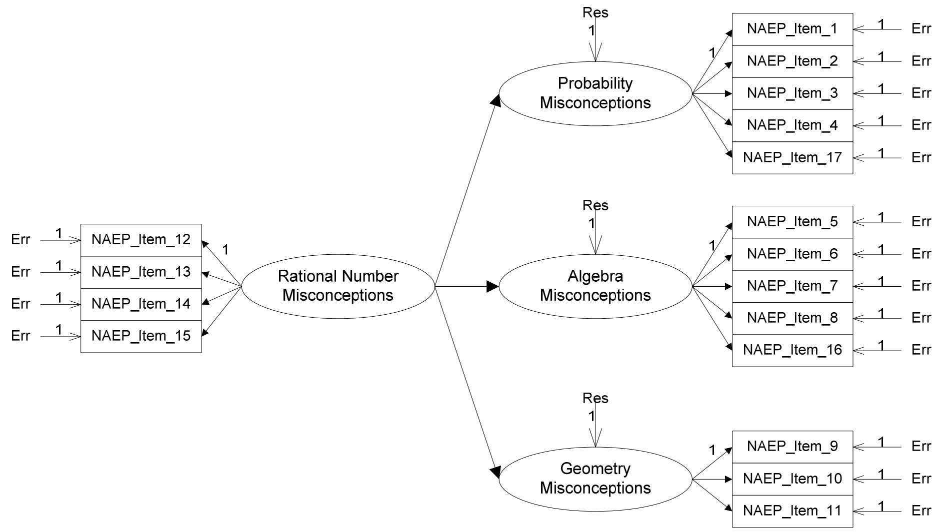
A


Err
NAEP_Item_1
NAEP_Item_2
NAEP_Item_3
NAEP_Item_4
NAEP_Item_17
Probability Misconceptions
Rational Number Misconceptions
Res
NAEP Item_12
NAEP Item_13
NAEP Item_14
NAEP Item_15
Algebra Misconceptions
Geometry Misconceptions
NAEP_Item_5
NAEP_Item_6
NAEP_Item_7
NAEP_Item_8
NAEP_Item_16
NAEP_Item_9
NAEP_Item_10
NAEP_Item_11
1

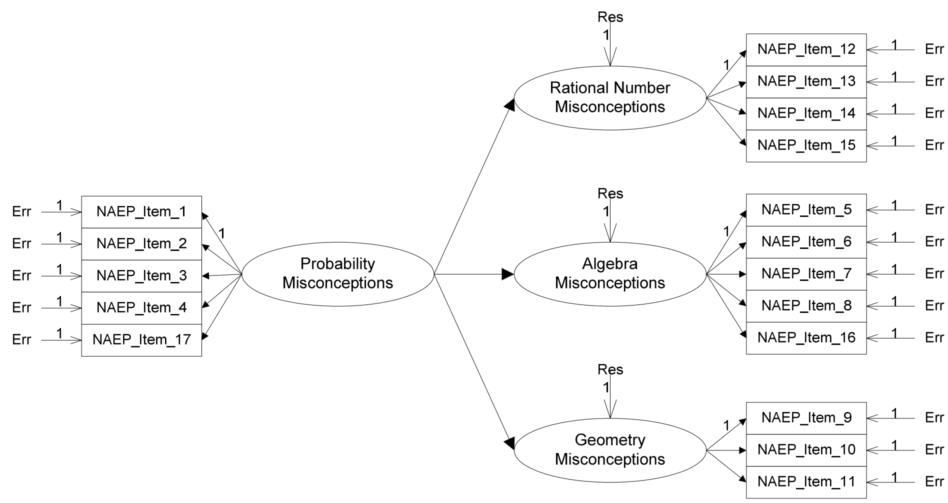
B


Err
NAEP Item_1
NAEP Item_2
NAEP Item_3
NAEP Item_4
NAEP Item_17
Probability Misconceptions
Rational Number Misconceptions
NAEP Item_12
NAEP Item_13
NAEP Item_14
NAEP Item_15
Res
Algebra Misconceptions
Geometry Misconceptions
NAEP_Item_5
NAEP_Item_6
NAEP_Item_7
NAEP_Item_8
NAEP_Item_16
NAEP_Item_9
NAEP_Item_10
NAEP_Item_11
1

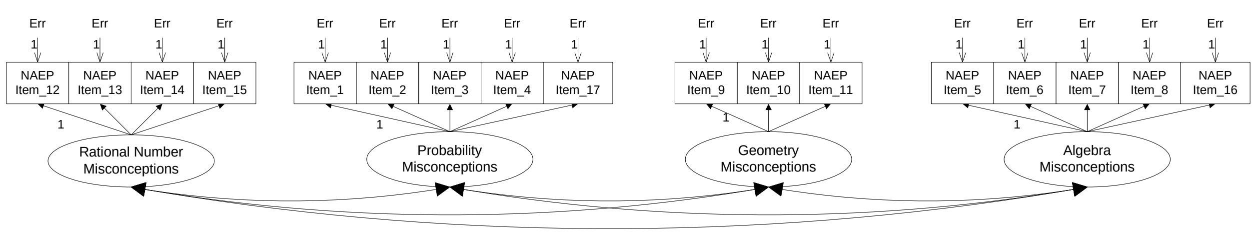
C



D



E



F

Figure 24. Hypothesized Structural Equation Models.

Model identification. Model identification is an extremely important initial consideration for any structural equation modeling (SEM) analysis and is determined by the degrees of freedom (*df*), computed by subtracting the number of freely estimated parameters from the number of independent data moments available (Byrne, 2009). If a model has less data moments than parameters (i.e., $df < 0$), it is considered *under-identified* (Byrne, 2009, p. 34). Because an infinite number of solutions are possible for an under-identified model, an analysis cannot proceed with such a relationship. If a model has an equal number of data moments and parameters (i.e., $df = 0$), it is considered *just-identified* (Byrne, 2009, p. 34). A just-identified model computes a unique solution to the model, but because there are no degrees of freedom, the model can never be rejected. The only acceptable model is one that is *over-identified* (Byrne, 2009, p. 34), having more data moments than parameters (i.e., $df > 0$). In an over-identified model, a unique solution can be computed, and that solution has a possibility of being rejected. In a multi-level structural equation model (i.e., all models except 12F), each level of the model must be over-identified as well as the full model. All six models in Figure 24 are over-identified along with each level within the model (Table 16).

Table 16
Degrees of Freedom for Models in Figure 24

Model	Full Model	Level 2	Level 3
A	116	14	4
B	116	9	9
C	114	53	—
D	116	14	—
E	116	20	—
F	113	—	—

Each multiple choice item was coded dichotomously as demonstrating a misconception or not based on the qualitative document analysis. Therefore, the polychoric and asymptotic covariance matrices were computed to adjust for the discontinuous nature of the observed variables (Byrne, 1998).

Goodness of fit indices were computed to provide supporting evidence for determining which model of misconception structure best matches the data. Since each fit index "operate[s] somewhat differently given the sample size, estimation procedure, model complexity, and/or violation of the underlying assumptions of multivariate normality and variable independence" (Byrne, 2001, p. 87), multiple fit indices were computed and compared to determine which model carries the least amount of misspecification for the data. The chi square statistic (χ^2) is the foundational statistic used to compare models (Hu & Bentler, 1995). χ^2 tests the null hypothesis that the covariance matrix reproduced according to the hypothesized model is the same as the population covariance matrix (Bandalos, 1993). However, as sample size increases, smaller differences will be magnified to the extent that unimportant differences will be statistically significant (Bandalos, 1993). This sensitivity to sample size led to the development of other goodness of fit statistics. Unfortunately the resulting statistics are also often prone to sample size correlation, sensitivity to non-normality, factor loading magnitudes, and model complexity. These statistics have been categorized as incremental, absolute, and residual-based absolute.

Incremental indices compare the target model test statistic to the independence model test statistic. Type 1 indices do so with no underlying distribution assumed, with the caveat that the same distribution is used for both the hypothesized and independence model. Type 2 indices rely on the central χ^2 distribution, a distribution with a mean of 0 and standard deviation of 1. Type 3 indices rely on the non-central χ^2 distribution. The present study will report the Comparative Fit Index (CFI), a Type 3 incremental index recommended by Byrne (2009), Hu and Bentler (1995), and Goffin (1993). Byrne (2009) recommended considering a CFI greater than 0.95 to reflect a well-fitting model.

Absolute fit indices approach the analysis of model fit from a different perspective than the incremental indices: rather than examining the hypothesized model against the independence model, these indices examine how the hypothesized model compares to the null model (i.e., not having a model). Marsh et al. (1988) found that the Goodness of Fit index (GFI; Jöreskog & Sörbom, 1984) provided the most accurate results from the absolute index category. For example, although the GFI is correlated with sample size, it does not inflate Type I error for sample sizes greater than 100 (Shevlin & Miles, 1998). Furthermore, its measurements are robust against latent dependence for sample sizes greater than 250 (Shevlin & Miles, 1998). The GFI is not, however, robust against non-normality at sample sizes below 500, and the present study will rely on dichotomous data, which are not normal. Since each half of the pretest sample will be greater than 500, this weakness in the GFI is not considered a major threat to validity. A GFI value greater than 0.90 represents a well-fitting model.

Additionally, Browne and Cudeck's (1989, 1993) Expected Cross Validation Index (ECVI) is an adjusted absolute fit index that measures a model's ability to hold in the population beyond a single sample by adjusting Akaike's Information Criterion (AIC; Akaike, 1973, 1983). Byrne (2009) recommended comparing ECVI values across models: The model with the smallest ECVI "exhibits the greatest potential for replication" (p. 82).

Residual-based absolute fit indices also compare the hypothesized model to the null model, but these statistics examine the residuals (i.e., unexplained variance) rather than the explained variance. The two primary residual fit indices are the Root Mean Square Error of Approximation (RMSEA; Steiger, 1998; Steiger & Lind, 1980; Steiger, Shapiro, & Browne, 1985) and the Root Mean Residual (RMR; Jöreskog & Sörbom, 1984). RMSEA has been found to be uncorrelated with sample size but moderately influenced by the magnitude of factor loadings and model complexity (Bandalos, 2009; Hu & Bentler, 1995). Byrne (2009) considered an RMSEA value less than 0.05 to represent a good-fitting model. RMR has been found to be highly sensitive to non-normality, but is less correlated with factor loading magnitudes and model complexity. RMR is unstandardized with a lower bound of zero and no upper bound. Interpretation is, therefore, problematic: RMR values must be compared to sample variance/covariance magnitudes. Jöreskog and Sörbom (1996) presented a modified version of RMR based on standardized values (i.e., correlations instead of covariances). The resultant statistic, the standardized RMR (SRMR) is bounded between zero and one. Kline (2005) found that SRMR values less than 0.10 represented good-fitting models.

A subcategory of indices, the parsimony-adjusted indices, does not measure goodness of fit; instead, these statistics penalize fit statistics for increasing model complexity. Including these parsimony-adjusted indices allows the researcher to simultaneously examine two interdependent pieces of information about a model: the goodness (or badness) of fit and "how parsimonious the model was in its use of the data in achieving that goodness of fit" (Mulaik et al., 1989, p. 439). In the present study, the parsimony version of the GFI and CFI (PGFI and PCFI) were provided to assess model complexity from the perspective of both incremental and absolute fit indices. Mulaik et al. (1989) and Byrne (2009) suggested that values of 0.5 or greater for both indices are not uncommon in acceptable models.

This mosaic of fit indices provides multiple perspectives of how well each model fits. The particular set of indices described above spans the various types of fit indices and a wide array of strengths and weaknesses. By incorporating all of them into the proposed analysis, confidence in the best fitting model will have a stronger foundation.

Analysis of contextual factors. In addition to the analysis of misconception structures, the present study also examined the impact of the probability intervention, attitudes toward mathematics, and metacognition on mathematical misconceptions using hierarchical generalized linear models (HGLM). The Bernoulli HGLM model best fits the dichotomous nature of the outcome data (Raudenbush & Bryk, 2002). In a Bernoulli model, the outcome variable is transformed to a logit, η, of the odds ratio, φ, for the outcome variable as seen in Equation 16. The computed logit becomes the linear outcome variable as seen in Equation 17. In this analysis, NAEP items (Level 1) are nested within students (Level 2, Equation 18), and students are nested within classes (Level 3, Equation 19). Variable abbreviations are defined in Table 17.

Hypothesized Item Level 1 Model

$$Prob\left(Misconception_{ijk} = 1 \mid \pi_{jk}\right) = \varphi_{ijk} \tag{15}$$

$$Log\left[\frac{\varphi_{ijk}}{1-\varphi_{ijk}}\right] = \eta_{ijk} \tag{16}$$

$$\begin{aligned}\eta_{ijk} = \pi_{0jk} &+ \pi_{1jk}\left(Item\ Discrimination_{ij}\right) \\ &+ \pi_{2jk}\left(Item\ Difficulty_{ij}\right) + \pi_{3jk}\left(Moderate\ Complexity_{ij}\right) + e_{ijk}\end{aligned} \tag{17}$$

Hypothesized Student Level 2 Initial Model (18)

For q = 0 to 3,

$$\begin{aligned}\pi_{qjk} = \beta_{q0k} &+ \beta_{q1k}\left(Conf_{jk} - \overline{Conf_{\bullet k}}\right) + \beta_{q2k}\left(Value_{jk} - \overline{Value_{\bullet k}}\right) \\ &+ \beta_{q3k}\left(Enjoy_{jk} - \overline{Enjoy_{\bullet k}}\right) + \beta_{q4k}\left(Mot_{jk} - \overline{Mot_{\bullet k}}\right) + \beta_{q5k}\left(KCog_{jk} - \overline{KCog_{\bullet k}}\right) \\ &+ \beta_{q6k}\left(RCog_{jk} - \overline{RCog_{\bullet k}}\right) + \beta_{q7k}\left(NAEP\_Pre_{jk} - \overline{NAEP\_Pre_{\bullet k}}\right) + r_{qjk}\end{aligned}$$

Hypothesized Class Level 3 Initial Model (19)

For q = 0 to 3, p = 0 to 7

$$
\begin{aligned}
\beta_{qpk} = \gamma_{qp0} + \gamma_{qp1}(Treatment_k) + \gamma_{qp2}\left(Mean\_Conf_k - \overline{Mean\_Conf_{\bullet}}\right) \\
+ \gamma_{qp3}\left(Mean\_Value_k - \overline{Mean\_Value_{\bullet}}\right) \\
+ \gamma_{qp4}\left(MeanEnjoy_k - \overline{Mean\_Enjoy_{\bullet}}\right) \\
+ \gamma_{qp5}\left(Mean\_Mot_k - \overline{Mean\_Mot_{\bullet}}\right) \\
+ \gamma_{qp6}\left(Mean\_KCog_{jk} - \overline{MeanKCog_{\bullet k}}\right) \\
+ \gamma_{qp7}\left(MeanRCog_{jk} - \overline{MeanRCog_{\bullet k}}\right) \\
+ \gamma_{qp8}\left(Mean\_NAEP\_Pre_{jk} - \overline{Mean\_NAEP\_Pre_{\bullet k}}\right) \\
+ u_{qpk}
\end{aligned}
$$

Table 17

Declaration of Variables in Equations 15 – 19

Variable	Description
Conf	ATMI Self Confidence subscale.
Enjoy	ATMI Enjoyment of Mathematics subscale.
Difficulty	Item Response Theory Parameter *b*.
Discrimination	Item Response Theory Parameter *a*.
Guessing/Chance	Item Response Theory Parameter *c*.
KCog	MAI Knowledge of Cognition subscale.
Mean_Conf	Classroom average of AMTI Self Confidence subscale.
Mean_Enjoy	Classroom average of ATMI Enjoyment of Mathematics subscale.
Mean_KCog	Classroom average of MAI Knowledge of Cognition subscale.
Mean_Mot	Classroom average of ATMI Motivation subscale.
Mean_NAEP_Pre	Classroom average NAEP mathematics knowledge pretest score.
Mean_RCog	Classroom average of MAI Regulation of Cognition subscale.
Mean_Value	Classroom average of ATMI Valuing Mathematics subscale.
Mot	ATMI Motivation subscale.
NAEP_Pre	NAEP mathematics knowledge pretest score
RCog	MAI Regulation of Cognition subscale.
Treatment	Indicator Variable for Treatment Group Assignment.
Value	ATMI Valuing Mathematics subscale.

Scale variables were centered to facilitate interpretation of the intercepts and slopes. In the student Level 2 equation (Equation 18), centering occurred at the group level (noted by the subscripts jk and $\bullet k$), causing the intercepts and slopes to be interpretable as student deviation from the classroom average (i.e., intercept represents the average student in the same class on all predictor variables; slope for any variable represents the impact of being an above or below average student in the class). In the classroom Level 3 equation (Equation 19), scale variables were grand mean centered, meaning that the overall mean is subtracted from each classroom mean (noted by the subscripts k and •). Grand mean centering changes the interpretation of the classroom Level 3 intercepts and slopes just as group centering did on the student Level 2 intercepts and slopes. At the classroom level, intercepts now represented the value for the average classroom on all predictor variables, and slopes represented the impact of being in an above or below average classroom. The regression coefficient for each classroom variable therefore represented the effect of a variable on the impact of each corresponding student variable. The variables e_{ijk} in Equation 17, r_{qjk} in Equation 18, and u_{qpk} in Equation 19 represented the random error measurement at the respective levels. These variables can be considered the random effect not captured by the model at each level, the unique effect of an individual item, student, or classroom on the effect, or variance not explained by the model at each level. The generic forms presented here include these random effects as potentially applicable to each equation. In reality, these random coefficients may or may not be desirable for a particular model. If present, the slope for any particular item, student, or school may vary uniquely within the respective group. If absent, the slope for all items, students, or schools on a particular equation are held constant. Therefore, an equation that excludes the random effect is used to answer questions about an effect controlling for a particular variable (i.e., excluding the effect of a particular variable); an equation that includes the random effect is used to answer questions about the effects of those variables (Raudenbush & Bryk, 2002). Equations that exclude random effects are especially useful for questions that involve an analysis of covariance (Keppel & Wickens, 2004). For example, a pre-posttest design usually involves questions about student growth during a treatment period. The effect of the pretest on the posttest is not interesting in and of itself; rather, removing that effect from the observed gain is critical to understanding the effect of the treatment. An equation designed to answer this type of question would therefore exclude the random coefficient.

Summary of Methodology

Five fundamental mathematics misconceptions were discovered through an examination of previous studies (e.g., Clement, 1982; Clements & Battista, 1992; Falk, 1992; Kahneman & Tversky, 1973a, 1973b, 1982, 1983; Küchemann, 1978; Shaughnessy & Bergman, 1993; Warren, 2000). Seventeen NAEP mathematics items were compiled into an instrument to test for these misconceptions within algebra, geometry, rational numbers, and probability. The difficulty, discrimination, and guessing coefficients for each NAEP item were measured using Item Response Theory. A unit of probability instruction was developed as a treatment for misconceptions based on the rationale that the abstract connections within probability's abundant concrete explorations and simulations would help students understanding the meaning of abstract concepts and the connections between ideas. A randomized pretest-posttest design with a switching replication was used to test this hypothesis. In addition to NAEP content area scores (percent correct and percent of misconception errors), two surveys were administered to measure the impact of contextual factors related to mathematics misconceptions. The Attitudes Toward Mathematics Inventory (ATMI) measured four factors of mathematics orientation (enjoyment, value, self confidence, and motivation; Tapia & Marsh, 2004); the Metacognitive Awareness Inventory (MAI) measured two factors of metacognition (knowledge of cognition and regulation of cognition; Schraw & Dennison, 1994).

Treatment fidelity was assessed from classroom observations and teacher interviews. Student explanations on the NAEP instrument were examined qualitatively to code content area misconception responses. The structural relationship of content area misconceptions was examined using structural equation modeling. The impact of item, student, and class characteristics on the emergence of misconception errors for a particular task was examined using three level hierarchical generalized linear modeling and two level hierarchical linear modeling. Chapter 4 presents the results of these analyses.

CHAPTER 4

RESULTS

This chapter describes the results of three analyses. First, student responses were examined to differentiate between reasoning or procedural errors and errors indicating an underlying misconception. Second, hypothesized relationships among content area misconceptions were examined using structural equation modeling. Third, the impact of item, student, and class characteristics was measured using three-level hierarchical generalized linear modeling and two level hierarchical linear modeling. Observations and teacher interviews were conducted to establish fidelity of treatment implementation for the third analysis.

Identifying Misconception Patterns

A sub-sample was chosen for a qualitative analysis of patterns of misconception responses on the NAEP-based mathematics knowledge test. On all items, students were asked to provide an explanation for their response choice. Approximately 74% of the sample left these explanations blank. To improve representativeness of the overall sample, the qualitative sub-sample was chosen using purposive stratification across classes; specifically, tests were chosen to be part of the sub-sample if they filled in the explanation section of the test for most items. Such a sampling technique produced a selection bias — students who completed their explanations were more likely to choose the correct answer, resulting in a reduced sample for each distractor to each item. To help manage this bias, the sampling procedure continued until all distractors for each item were represented (N = 72). Division of items by content area (i.e., algebra, geometry, probability, and rational number) transferred directly from the NAEP classification of each item.

The following descriptions focus primarily on student explanations of errors; however, the analysis began with a recognition that correct responses do not necessarily indicate conceptual understanding. For all items in which the correct response explanations are not discussed explicitly, the explanations by students indicated that they did, in fact, understand the concept not understood by students who chose incorrect responses. The thick description provided in this analysis was used to establish trustworthiness for the coding of misconception responses.

Misconceptions on Algebra Content Knowledge Items

Algebra items (i.e., Items 5, 6, 7, 8, and 16; Appendix N) included distractors that reflected misconceptions about additive/multiplicative structures, the meaning and interpretation of variables, and the meaning of rational numbers.

Item 5 response patterns. Item 5 described the formula to convert temperature from Fahrenheit to Celsius in words and then asked students to convert 393°F to Celsius. I hypothesized that choices A, B, and C would represent misconceptions about the meaning of rational numbers. Student responses confirmed this hypothesis. For example, one student chose A, “Because 393 – 32 = 361; 5/9 = .55, so you divide 361 by 5/9, answer is 656.3, to the nearest degree is 650.” This explanation represents the explanations of others who chose A, indicating that students who chose A did so because they divided by the rational number rather than multiplying, not realizing that the resultant rational number, 5/9 of 361, should be smaller than 361.

I hypothesized that Choice B for Item 5 would result from the ignoring of the denominator, and again, student explanations confirmed this hypothesis. The most explicit case of this type chose B, “Because 393 x 5 = 1805.” This student also failed to note that he/she had subtracted the 32 from the 393 properly; 361 x 5 is 1805 while 393 x 5 is actually 1965. I decided that this particular error was simply one of reporting rather than a misconception, so it was excluded for the purposes of this analysis. Another student who chose B stated, “361 • 5/9 = 200 5/9 = 200 x 9 = 1800 + 5 = 1805.” This student failed to realize that the rational number was accounted for by the 200 and continued to try to incorporate the fraction, ultimately doing so by misusing both numbers. From this question, I considered how the student had correctly computed the 200 if he/she did not understand how to use the 5/9 later. My best guess was that the student used a calculator for the first computation, but thought that 200 x 9 would be an easy calculation, so he/she did the last steps by hand and did not check them on the calculator. Although this conclusion is wholly speculative, if true, it may suggest that the use of calculators to explore the meaning of rational numbers may open an avenue for addressing student conceptions and perceptions of rational numbers.

I also hypothesized that Choice C would represent a misconception about rational numbers, specifically, that students would choose C by ignoring the rational number altogether. Student responses also verified this hypothesis. Students who chose C justified their response with statements such as, “Divide 393 and 32.”

Originally, I hypothesized that E would not represent a similar misconception as those for A, B, and C on Item 5. Student responses, however, contradicted this hypothesis. Students who chose E also ignored the denominator and misused the numerator as did students who chose B. For example, students justified choice E with statements such as, “Divide the numbers,” specifically 361 by 5. Therefore, Choice E was added to the misconception choices for Item 5.

These interpretations of rational numbers appeared to typify student responses to rational numbers. I generalized these patterns of rational number interpretation into five types:

1. Rational number is understood to be a single quantity, but confusion about the meaning of that quantity results in the application of the wrong operation or the correct operation(s) to the wrong quantities (e.g., Divide instead of multiply, multiply by the wrong number). This error connects to rational number meaning misconceptions identified by Fosnot and Dolk (2002).
2. Reverse the role of the numerator and/or denominator. This error is similar to those described by Baturo (1994), Behr et al. (1992), and Lamon (1999).
3. Ignore either the numerator or denominator (as in Green, 1983b; Watson & Shaugnessy, 2004).
4. Ignore the numerator/denominator AND reverse the role of the remaining part of the rational number (e.g., Divide by the numerator)
5. Ignore the rational number altogether. This error appeared to connect to variable number misconception described by Küchemann (1978) in which students ignored the presence of variables.

While this categorization of rational number meaning errors may not account for every rational number meaning error for every problem, it may serve as a foundation for exploring rational number meaning errors in other problems/contexts. Additionally, this list appears to be hierarchical; that is, a Type 2 rational number misinterpretation may represent a greater degree of confusion about the meaning of rational numbers than Type 1, as would Types 3, 4, and 5 over a Type 2.

Item 6 response patterns. Item 6 offered students a linear function ($y = 4x$) and asked about the change in y based on an increase of two units to x. Prior to the present analysis, only Choices D and E were hypothesized as misconception responses — I expected students with additive/multiplicative structure misconceptions to choose D by squaring the independent variable coefficient and E by doubling that same coefficient (as described by Warren, 2000). Explanations from students who chose these two responses supported this expectation. For example, students who chose E typically showed their calculation as "4 x 2 = 8." Alternatively, students who chose A, the correct answer, also performed this same calculation but knew to add it to the overall y value rather than making it a new coefficient. Therefore, D and E remained misconception responses in the coding procedures. Additionally, explanations for choices B and C also indicated misconceptions in student reasoning about additive and multiplicative structures. For example, students who chose B stated, "4 + 2 = 6" or "It also increases by 2," similar to patterns found by Warren (2000) and Moss et al. (2008). Likewise, students who chose C offered one of three justifications, all of which represented a misconception about how to handle an additive structure in an algebraic equation. The first type of explanation demonstrated a reliance on the balance-beam principle of algebraic equations, stating something such as, "You basically add 2 more to the other side," "Because both sides need to be the same, or "Because if one increases, so does the other." The second type of explanation showed that some students chose C because they thought that the change of two should be added to the coefficient, stating rationales such as, "Because 4 + 2 = 6 that is 2 more than the original amount" or "Because 4x + 2 = 6x." The third type of response to support Choice C indicated that students knew that the change in *y* should be additive, but they failed to understand the role of the coefficient in that change. These students justified their choice of C by asserting, "It would be y = 4*x* + 2." As a result of this analysis, the coding of misconception responses was expanded to include choices B and C. These errors appeared to occur because students were relying on procedures isolated from meaning and connections between ideas. The framework in Figure 10 may shed light on how these errors emerged. For example, students who relied on the balance beam principle for solving algebra equations did not seem to understand why such an approach works and what it means to a particular context. These students appeared to demonstrate procedural knowledge with instrumental understanding (Skemp, 1976/2006). As a result, these students developed algebraic misconceptions about how, when, and why the algebra balance beam works. These misconceptions about the balance beam led to faulty reasoning that may have reinforced the balance beam misconceptions.

Item 7 response patterns. Item 7 asked students to choose an expression to represent the situation, "A plumber charges $48 for each hour and an additional $9 for travel." The correct response, choice E, uses the $48 per-hour charge as the coefficient to the number of hours and adds the $9 travel fee as a one-time charge. Every distractor response was hypothesized to represent a misconception about additive/multiplicative structures. Student explanations verified this prediction. For example, students who chose A interpreted each charge as an "additional" charge. Likewise, students who chose B thought the calculation resulting from such an expression should be "added on to the original." They also believed that both charges should be multiplied by the number of hours. Students who chose C knew that a charge should be added and another multiplied, but they reversed the quantities. Finally, students who chose D understood that the expression should represent “48 times the hours plus 9,” but they did not understand how to translate those words into an expression. I concluded that choice D could possibly represent variable misconceptions as well as additive/multiplicative structure misconceptions, but since the analyses of the present study will not differentiate between types of misconceptions, D was left as a misconception response.

Item 8 response patterns. Item 8 presented students with the following scenario: “Carmen sold 3 times as many hot dogs as Shawn. The two of them sold 152 hot dogs altogether. How many hot dogs did Carmen sell?” Originally, I hypothesized that choice B would represent the reversal error, reflecting the wrong person’s amount, similar to the reversal error identified by Clement (1982). The present analysis revealed that such was not the case in this sample. In fact, students who chose B provided correct equations such as “I did 3s + s = 152 and add the s to the 3s to get 4s and divided 152 and 4 by 4 and got s = 38.” Not one student who chose B in the sub-sample indicated that they thought the 38 represented Carmen’s amount. As a result, I concluded that choosing B did not represent a variable meaning misconception so much as a careless error; specifically, not catching that the question asked for Carmen’s amount instead of Shawn’s. Almost twice as many students in the sub-sample chose C instead of B, and these students did indicate one of Küchemann’s (1978) variable interpretation errors — they unanimously ignored the variables altogether and simply divided 152 by 3. Students who chose E also ignored the variables entirely, showing a similar calculation to that of C. For example, students who chose C justified their answers with statements such as, “Because 152 ÷ 3 gives you 50.8, round and you get 51.” Likewise, students who chose E wrote statements such as, “Because 50 + 50 + 50 = 150 – 2.” In both cases, students failed to recognize the role of the variable in partitioning the total amount. Other students who chose C and E demonstrated Küchemann's Level 1 interpretation, evaluating the variable using trial and error. Therefore, I concluded that choices C and E should represent the variable misconception rather than the original choice B.

Item 16 response patterns. Item 16 presented students with a table of values and asked them to determine the function that best modeled the data. I hypothesized that choice A would represent a misconception about the nature of the functional relationship; student responses verified this expectation. Students who chose A made statements that indicated an understanding of a relationship, but they looked at the relationship backwards, i.e., they thought of *n* as the dependent variable rather than the independent variable, similar to the reversal error in Clement (1982). Although such an inversion might also be due to a rational number meaning misconception (i.e., doubling rather than halving, as in Item 5, error 2, Baturo, 1994; Behr et al., 1992; Lamon, 1999), the purpose of the present analysis is not to distinguish between misconception types but rather to reflect the presence of misconceptions in a content area. Therefore, choice A remained a misconception response in the analysis. This overlap indicates the possibility of an underlying multicollinearity across content area misconceptions resulting from the influence of the underlying misconceptions.

Student explanations for choosing C or D on Item 16 also indicated the presence of variable misconceptions. Students who chose C explained that they had only used the first column of values to determine the equation, concluding that, "You subtract by the *n* and that equals *p*." For choice D, students explained that they thought the number of days was one of the variable quantities of importance, "Because the days matter too." These students then used the first column of values to conclude that subtraction held the key to solving this problem. As a result of this analysis, choices C and D were included as misconception responses for Item 16.

Misconceptions on Geometry Content Knowledge Items

The geometry items on the NAEP instrument (i.e., Items 9, 10, and 11; Appendix N) examined student misconceptions about spatial reasoning and the meaning of rational numbers.

Item 9 response patterns. Item 9 presented students with a rectangle and asked them whether the figure should be classified as a parallelogram. Students who chose the correct response, A, did so because, "It has parallel sides" or "It has equal sides." This explanation indicated that these students were operating at least at Van Hiele Level 1, in which students recognize that figures have characteristics and properties. Students who chose B, on the other hand, indicated operating at Van Hiele Level 0, in which students rely on visual recognition of shapes. For example, students made statements such as, "Parallelograms are crooked " and "The figure she drew has right angles." Other students indicated that they thought that being a rectangle and square excludes a shape from being a parallelogram through statements such as, "It is a square" and "No. It is a rectangle." These types of errors indicate a fundamental spatial reasoning misconception resulting from low Van Hiele levels of understanding (Clements & Battista, 1992; Crowley, 1987). Based on this evidence, I retained choice B as a misconception response.

Item 10 response patterns. Item 10 presented students with a shaded figure within a grid of centimeter squares. I hypothesized that Choice D would represent a spatial reasoning misconception, in which students would rely on the lengths to compute the area rather than on the meaning of area, similar to Clements and Battista's (1992) 9^{th} and 11^{th} most common spatial reasoning misconceptions. Explanations by students who chose D verified this hypothesis with statements such as, "Because you multiply the length and width of squares." Additionally, to arrive at Choice D, students also needed to be confused about the length of the diagonals; in this case, they evidently chose to add them as a little more than 1, then rounded the length to an even 7 cm. They also failed to recognize that the width on one end was 2 cm while on the other end it was 3 cm. Therefore, this choice was retained as a misconception choice.

Additionally, an analysis of Choice C explanations revealed the presence of additive/multiplicative structure and spatial reasoning misconceptions, or Van Dooren et al.'s (2003) illusion of linearity applied to geometric shapes. Students who chose C explained that they simply "counted them all up," approximating the diagonal lengths as 1.5 cm and computing a perimeter rather than an area. Therefore, I added Choice C as a misconception choice for Item 10.

Students who chose A for their response to Item 10 also revealed an error in spatial reasoning. These students recognized that they needed to count shaded areas, but they counted only the wholly shaded squares and ignored the half shaded squares, as indicated by statements such as, "There are 9 squares that are fully shaded" and "The image is mainly 3 by 3 so 9 sq centimeters, a = 3 • 3 = 9." These explanations led me to believe that Choice A resulted from faulty reasoning, but not necessarily a misconception — they knew to add areas, and they appeared to recognize that area is the space inside a closed figure. Therefore, Choice A was not added to the misconception choices for Item 10.

Item 11 response patterns. For Item 11, students were asked to convert the dimensions of an object from one unit of measure to another. I hypothesized that the difficulty in Item 11 for students lay in the recognition that the dimensions represented two quantities rather than one, so the misconception of interest was the way students ignored these variable quantities. Some students ignored one of the variables; others thought that the variables were simply labels and therefore traded labels in a one-to-one relationship. Prior to this investigation, I hypothesized only that choice A would represent a misconception response, indicating the use of the variable as a label. Explanations by students who chose A confirmed this hypothesis, indicating that the values "don't change."

Additionally, students who chose B and C also indicated that only one variable needed to be accounted for in the computation. Students justified their choices with explanations such as, "I did 5 x 3 = 15, then left 3 the same." Therefore, choices B and C were added to the list of misconception responses for Item 11.

From this investigation, misconception responses for geometry items were expanded to include additional choices. Because misconceptions such as additive/multiplicative structure were also evident in algebra items, items from both content domains may co-vary in the SEM analysis.

Misconceptions on Probability Content Knowledge Items

Probability items (i.e., Items 1, 2, 3, 4, and 17; Appendix N) on the NAEP instrument examined student errors in probability rooted in absolute/relative comparison, additive/multiplicative structure, spatial reasoning, and rational number meaning misconceptions.

Item 1 response patterns. For Item 1, students were asked to determine which picture represented the greatest probability. I hypothesized that misconception responses would follow the patterns identified by Shaughnessy & Bergman (1993) and Watson & Shaughnessy (2004): Students with probability misconceptions would focus on the number of black marbles rather than the ratio of black to white marbles, or, confusing absolute and relative comparisons. Students demonstrated this misconception precisely as expected. For example, students who chose answer C stated that its dish “contains more blacks” or “has the most black marbles.” Likewise, students who chose E also indicated that the number of black marbles was the only number of importance. Students who chose “B” made two absolute comparisons, looking for a combination of the “most white and black.”

Item 2 response patterns. Item 2 presented students with a two-way table (gender by color of puppies) and asked students to compute the conditional probability of a puppy being male given that it is brown. I hypothesized that Choices B, C, D, and E would indicate a misconception about the meaning of rational numbers, as described by Bar-Hillel and Falk (1982) and Falk (1992). Student explanations for these responses verified this hypothesis.

Students who chose B on Item 2 ignored the condition of being brown altogether and instead gave the probability of being male, providing explanations such as “Because it’s seven puppies and two of them are male” or “2 total males and 7 total puppies; chance a male will be picked 2/7.” These students appeared to be unsure of how to incorporate the brown condition into the probability quantity.

Students who chose C and E on Item 2 gave explanations that indicated confusion between part-part relationships (i.e., odds) and part-whole relationships (i.e., probability). These students justified their choices with statements such as, “There’s 1 male and 3 girls so the probability is 1/3” or “Because there’s 2 female and 3 male.” One student who chose C, however, did so because, “There are 3 black puppies and 1 is a male.” This student, rather than being confused about the meaning of rational numbers, simply computed the wrong conditional probability (black instead of brown), and he/she did so correctly. This way of choosing C appeared to be an abberation rather than the pattern, so although C could be reached through a reasoning process not emerging from a misconception, the explanations of students in this sample indicated that the choice was overwhelmingly due to the misconception. No such abberations appeared in explanations for Choice E. Therefore, both C and E were retained as a misconception response.

Students who chose D for Item 2 provided explanations that indicated two reasoning processes, both of which represented thinking based on misconceptions about rational numbers. The first explanation, used by the majority of students who chose D, relied on a comparison of brown dogs to dogs; these students made statements such as, "There are 2 male puppies and only 1 is brown" or "Because all together there is 1 male black and 1 brown, add them up which = 2 (so, 1/2)." These responses indicated confusion about which quantities should have been represented by the part whole relationship. The second type of explanation relied instead on the uniformity heuristic as described by Falk (1992). These students chose D, "Because there are only 2 types of genders that you can pick." Whether from confusion about part-whole relationship quantities or the uniformity heuristic, the evidence from student explanations indicated that choice D did represent a misconception.

Item 3 response patterns. Item 3 also asked students about a conditional probability. In this scenario, Bill had a bag of 30 candies, 10 each of red, blue, and green. Using a random draw, Bill ate two pieces of blue candy. Students are then asked if the probability of getting a blue candy on the next draw is still 10/30 or 1/3. I hypothesized that students who chose A (yes) would ignore the conditional aspect of the probability altogether and that students who chose B (no) would recognize that the quantity making up the part-whole relationship had changed. Student explanations to both choices verified this hypothesis. Students who chose A sometimes relied on the uniformity heursistic, making statements such as, “Because there are 3 colors, and 1 could be picked blue.” Others who chose A simply ignored the conditional, justifying their response with statements such as, “You have 10 of each color candy to add up to 10/30” or “Because there are ten and all together are 30.”

In contrast to explanations of choice A, students who chose B did not appear to do so by guessing or elimination. Indeed, these students recognized that the consumption of the two pieces of blue candy changed the quantities represented by the probability: “He ate 2/10, so it’s now 8 blues instead of 10,” “Because he has already eaten 2, which lessens his chances,” “He already ate 2 of them, so its 28 left,” “Because his chances go down,” or “Because he took out 2 candies; it’s now 8/28.” Based on these explanations, I concluded that students in this sample who chose correctly did indeed demonstrate a stronger conceptual understanding of the meaning of the rational number quantities present in the probability, so Choice A was retained as a misconception indicator for Item 3.

Item 4 response patterns. Item 4 presented students with a spinner divided in two halves, with one of the halves divided in half again. The arrow on the spinner pointed to one of the quarter regions. Students were then asked how many times they should expect the arrow to land in that region after 300 spins. I hypothesized that a spatial reasoning misconception would result in students deciding that the probability of the region was 1/3 instead of 1/4. Based on this hypothesis, I expected Choice C to result from students taking 1/3 of 300. I also expected students to choose D by taking 1/3 of 360, the number of degrees of the circle. Explanations for choosing C verified that part of my hypothesis. These students made statements such as, “Because there’s 3 spaces and 300 ÷ 3 is 100” or “I divided 300 ÷ 3 & got 100.” Therefore, I retained Choice C as a misconception choice. Students who chose D, on the other hand, did not verify my hypothesis. Instead, these students indicated that they had recognized the probability of the region as being 1/4 but made a computation error. Noone who chose D gave explanations that indicated a misconception, so it was eliminated as a misconception choice. Explanations for choosing B did, however, indicate a misconception, possibly one of spatial reasoning or rational number meaning. These students recognized that the region was 1/4 of the circle, so they fell back on the number of degrees in a circle. These students made statements such as, “The circle is in an angle of 90º,” “Circle = 360, divide it by 4, you get 90,” or “Because the circle is split up into 3 parts; a circle’s measure is 360, if cut in half, each part will be 180, if one half of the split circle is cut in half again, that side is now 2 sets of 90º.” This error could be due to misunderstanding about the meaning of the quantities in a probability ratio, or it could be due to misunderstanding the quantity being predicted, focusing on a single circle instead of the same circle 300 times. Regardless of which misconception led to this choice, it seemed clear that choosing B for Item 4 represented at least one type of misconception, so it was added as a misconception choice for this Item. As with Item 16, the convergence of fundamental misconceptions on multiple content areas increased the likelihood of a high degree of collinearity between content area misconceptions.

Item 17 response patterns. Item 17 asked students to visualize a cube whose faces are labeled R or S. The probability of landing on R was given as 1/3. Students were then asked to determine how many faces of the cube should be labeled R. I hypothesized that Choices C and E would represent misconceptions about absolute/relative comparisons or the meaning of rational numbers. Students who chose these responses and explained their answers corroborated this expectation with statements such as "because 1/3 → 1," "one because there is a one out of three chance," or "the number on the bottom is how many." Some students who chose C ignored the stated probability altogether, similar to students who ignored rational number quantities in Item 5 (as students did with variables in Küchemann, 1978). These students relied instead on the uniformity heuristic (Falk, 1992), stating that "each face R and S gets three sides" and "half of 6 is 3." From these explanations, I concluded that C and E should be retained as representative of misconception-based reasoning.

Misconceptions on Rational Number Content Knowledge Items

Rational number items (i.e., Items 12, 13, 14, and 15; Appendix N) on the NAEP instrument examined student errors on rational numbers based on absolute/relative comparison, additive/multiplicative structure, and rational number meaning misconceptions.

Item 12 response patterns. Item 12 presented students with a situation in which the postage cost for a letter is based on a different rate for the first ounce. I hypothesized that Choice D would indicate an additive/multiplicative structure misconception (as in Moss et al., 2008; Warren, 2000). Explanations by students who chose D confirmed this hypothesis, making statements such as, "I multiplied .33¢ times 2.7" and "Because 33 + 33 = 66; 0.7 → 1 ➔ 66 + 22 = 88." Student explanations of other choices revealed additional misconception responses for Item 12. Students who chose E followed the same reasoning as students who chose D, but they remembered to round the 2.7 to 3. So, these students wrote explanations such as "33 x 3 ounces = 99 cents." Students who chose B used two types of reasoning to arrive at their answer. First, students added "33 + 22 + 11," "33 + 22 + 0.7," or "You have 2 whole ounces, 33¢ + 22¢ = 55¢, next you have to figure out the .7. Take 22 • 70%, which is 18. So you add 12 to 55, total would be 66¢." This final statement, apart from the readily apparent calculation errors, shows the same basic reasoning as the first two. This particular justification included several erasures over numbers, a characteristic that appeared to indicate that the student had changed numbers to arrive at the closest answer available. So, not only does the response demonstrate the same additive/multiplicative structure misconception, it may also indicate the persistence of the misconception even in the face of numbers not adding up correctly. The second type of justification for choice B relied on a multiplicative-only strategy, for example, "I multiplied 33 times 2." Students who chose A used similar reasoning to that used by students who chose B. These students also dropped the 0.7 or, in one case, the second ounce. Most of these justifications were some variation of, "33 + 22 = 55." The student who ignored the second ounce stated, "First ounce is 33 cents, next 0.7 of ounce is 22, rounded." As a result of this analysis, I included choices A, B, and E as misconception responses in addition to the original hypothesized choice D. The presence of additive/multiplicative structure misconceptions in rational numbers as well as algebra emphasizes the likelihood of collinearity between content area misconceptions.

Item 13 response patterns. Item 13 presented students with a diagram and a scale of 3/4 in = 10 ft. They were then asked to convert 48 feet to the scale drawing length. I hypothesized that Choices D and E would represent reasoning indicative of rational number meaning misconceptions. Students chose D because, "3/4 x 10 = 7.5 in" or "3 ÷ 4 = 0.75, 7.5 in." These students demonstrated the first rational number misinterpretation identified by the analysis of Item 5 responses – performing the correct rational number operation to the wrong quantity (as in Fosnot & Dolk, 2002). Therefore, Choice D was retained as a misconception response.

Students who chose E, however, did not justify their choices with responses that implied that they had ignored the numerator as I had hypothesized. Instead, they made statements such as, “Because you add 3/4 to 48 ft” or “Add them all up and divide by 10.” These responses clearly indicated an error in thinking, but I chose to discard E as a misconception choice because I could not find a clear connection between misconceptions identified by previous research and the reasoning process that led to this choice.

Students who chose C for Item 13 indicated a misconception that I did not anticipate in my hypotheses. These students indicated that they had ignored the rational number altogether, the fifth rational number mis-interpretation identified by the analysis of Item 5. Some students justified their choice with statements such as, “48 ÷ 10 = 4.8 or 5.” Therefore, choice C was included as a misconception response for Item 13.

Item 14 response patterns. In Item 14, students were asked to arrange a set of three fractions in ascending order, distinguishing between absolute and relative comparisons. Students who responded with choice E recognized the relative relationship between fractions, but they misunderstood the nature of that relationship. For example, students understood that fractional numbers compare differently than other rational number forms, but they failed to recognize that the denominator quantity is the one that has this inverse relationship (as in Baturo, 1994; Behr et al., 1992; Lamon, 1999). Students who chose E simply stated that “smaller fractions are larger.” Although this faulty reasoning clearly demonstrates a misunderstanding, these students did not demonstrate the absolute/relative comparison misconception, so E was not included as a misconception response. In constrast, students who chose B cited a comparison of the numerators only, ignoring the impact of the denominator on the overall quantity (i.e., absolute versus relative comparison misconception). Therefore, B was the only response recognized as a misconception response for Item 14.

Item 15 response patterns. Item 15 presented students with a rectangle divided into two columns of 10 cells for a total of 20 cells, six of which were shaded. Students were then asked to determine which rational number best represented the probability of the shaded region. I hypothesized that students who chose D would do so because they used a part-part relationship rather than a part-whole relationship. Explanations by students who chose D verified this hypothesis, making statements such as, "3 are black and 7 are white." Therefore, Choice D was retained as a misconception choice for Item 15.

Implications of Item Response Patterns

This analysis of student responses to the NAEP mathematics knowledge test fundamentally altered the way misconception responses were coded. For several items, student responses validated the hypothesized misconception choices and rationales for each choice. For several other items, student responses suggested that the hypothesized misconception responses were either not due to misconceptions at all, not due to the hypothesized misconception, or not due to the hypothesized misconception for the correct reasons. As a result, the coding of these items was changed to match student response patterns to maximize content validity prior to the structural analysis of content area misconceptions and the analysis of contextual factors on the emergence of misconceptions on a particular item. Table 18 summarizes the changes from hypothesized to observed misconception responses.

Table 18
Summary of Observed Misconception Responses

Item	Hypothesized Misconception Responses	Observed Misconception Responses
1	C, E	C, E
2	B, C, D, E	B, C, D, E
3	A	A
4*	C,D	B, C
5*	A, B, C	A, B, C, E
6*	D, E	B, C, D, E
7	A, B, C, D	A, B, C, D
8*	B	C, E
9	B	B
10*	D	C, D
11*	A	A, B, C
12*	D	A, B, D, E
13*	D, E	C, D, E
14	B	B
15	D	D
16*	A	A, C, D
17	C, E	C, E

*Indicates an item where coding was changed because of response analysis.

Structure of Content Area Misconceptions

To examine the structure of misconceptions between content areas in secondary mathematics, six potential theoretical models were hypothesized (Figure 24). These models were compared by their goodness of fit indices. Additionally, the six data sets were randomly split into two data files each, one for calibration of the structural models and the other for validation of the resultant models (as recommended by Byrne, 1998) to maximize convergent validity.

Calibration of Hypothesized Structural Models

Patterns in the modification indices (MIs) across imputations were examined to ensure that the maximum amount of error within the model had been explained. I focused on the maximum modification index from each data set; however, I also looked for high MIs that matched the maximum MI from other data sets for the same model. Instead of finding high matches, I found that the maximum modification index in one data set was invariably small in the others (i.e., less than 4, $p > 0.05$). Therefore, I focused instead on the theoretical relevance of each maximum modification index. I considered an MI to be theoretically sound based on the content knowledge examined and the underlying misconception addressed.

Analysis of Model A

Model A specified misconceptions in rational number content area as an independent variable that predicts misconceptions in probability content area, which in turn predicted misconceptions in algebra and geometry (Figure 24A). Therefore, in this model, probability acted as a filter for rational number. The goodness of fit indices suggested that this model fit the data very well (Table 19).

Table 19
Model A Goodness of Fit Indices from Calibration Samples

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	606.76	0.99	0.87	0.024	[0.009, 0.035]	0.075	0.66	0.728	1.42
1 (N = 577)	716.34	0.98	0.87	0.031	[0.022, 0.040]	0.076	0.66	0.714	1.37
2 (N = 553)	680.20	0.99	0.89	0.018	[0.000, 0.029]	0.066	0.68	0.726	1.15
3 (N = 558)	778.03	0.98	0.87	0.032	[0.023, 0.041]	0.076	0.66	0.721	1.44
4 (N = 566)	646.52	0.99	0.88	0.025	[0.014, 0.034]	0.074	0.67	0.716	1.25
5 (N = 575)	716.60	0.99	0.88	0.025	[0.013, 0.034]	0.072	0.67	0.723	1.25
Wtd. Avg.	707.53	0.99	0.88	0.026	[0.017, 0.035]	0.073	0.68	0.748	1.29
SE_{Avg}	24.494	0.003	0.004	0.003		0.002	0.004	0.004	0.057

Note: *df* = 116

Calibration. Model A MIs most commonly called for covariances between Item 6 and 12 and Items 5 and 15 error terms (Table 20). Items 6 and 12 both measured additive/multiplicative structure misconceptions, so a relationship between these two variables seemed plausible even though they measured different content knowledge. Likewise, Items 5 and 15 both measured rational number meaning misconceptions.

Table 20
Model A Maximum Modification Indices from Calibration Samples

Imputation	Maximum Modification Index	Associated Parameter to Add
0	33.17	TH (4, 6); Covariance between Item 5 and 15 errors
1	339.07	TH (1, 8); Covariance between Item 7 and 12 errors
2	43.79	TH (1, 7); Covariance between Item 6 and 12 errors
3	894.40	TH (1, 7); Covariance between Item 6 and 12 errors
4	68.22	TD (3, 2); Covariance between Item 13 and 14 errors
5	361.69	TH (4, 6); Covariance between Item 5 and 15 errors

Since the unimputed data file called for the error covariance of Items 5 and 15, this parameter was added to the model and the goodness of fit statistics were computed (Table 21).

Table 21
Model A2 Goodness of Fit Indices, Covary Item 5 and 15 Errors

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	578.83	0.99	0.88	0.021	[0.000, 0.033]	0.075	0.66	0.743	1.36
1 (N = 577)	750.91	0.97	0.87	0.032	[0.023, 0.040]	0.075	0.66	0.736	1.37
2 (N = 553)	679.00	0.99	0.89	0.019	[0.000, 0.030]	0.066	0.67	0.745	1.15
3 (N = 558)	769.37	0.98	0.87	0.031	[0.022, 0.040]	0.075	0.65	0.732	1.40
4 (N = 566)	627.74	0.99	0.89	0.022	[0.009, 0.032]	0.073	0.67	0.745	1.19
5 (N = 575)	680.75	0.99	0.89	0.021	[0.0073, 0.031]	0.070	0.67	0.745	1.19
Wtd. Avg.	701.59	0.98	0.88	0.025	[0.015, 0.035]	0.072	0.664	0.741	1.26
SE_{Avg}	28.967	0.004	0.005	0.003		0.002	0.004	0.003	0.058

Note: *df* = 115

The difference of χ^2 test revealed a statistically significant reduction of model misfit for both the unimputed data file (listwise deletion) and the imputed data files ($\Delta\chi^2_{Unimputed}[1] = 27.93$, $p < 0.0001$; $\Delta\chi^2_{Imputed\ Avg}[1] = 5.935$, $p = 0.015$). I therefore, retained Model A2 and examined its modification indices (Table 22).

Table 22
Model A2 Maximum Modification Indices

Imputation	Maximum Modification Index	Associated Parameter to Add
0	311.76	BE (2, 3); Regression path from Geometry to Algebra
1	178.20	TH (1, 8); Covariance between Item 7 and 12 errors
2	60.24	TD (4, 1); Covariance between Item 12 and 15 errors
3	64.18	TE (4, 1); Covariance between Item 1 and 4 errors
4	66.25	TD (3, 2); Covariance between Item 13 and 14 errors
5	39.05	TE (12, 11); Covariance between Item 9 and 10 errors

The addition of the regression path from geometry to algebra created an unstable model that would not converge across the data sets. The covariance between Items 7 and

12 indicated by Imputation 1 made theoretical sense to me since both items measured additive/multiplicative structure misconceptions. So, the parameter was added to the model, and the goodness of fit statistics were computed (Table 23).

Table 23
Model A3 Goodness of Fit Indices, Covary Item 7 and 12 Errors

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	578.82	0.99	0.88	0.022	[0.000, 0.033]	0.075	0.65	0.657	1.36
1 (N = 577)	743.84	0.98	0.87	0.032	[0.023, 0.040]	0.075	0.65	0.657	1.36
2 (N = 553)	675.08	0.99	0.90	0.019	[0.000, 0.030]	0.066	0.67	0.663	1.13
3 (N = 558)	769.14	0.98	0.87	0.031	[0.022, 0.040]	0.075	0.65	0.635	1.40
4 (N = 566)	627.27	0.99	0.89	0.023	[0.0097, 0.033]	0.073	0.66	0.660	1.19
5 (N = 575)	677.92	0.99	0.89	0.022	[0.0077, 0.032]	0.070	0.66	0.667	1.19
Wtd. Avg.	698.67	0.99	0.88	0.025	[0.016, 0.035]	0.072	0.66	0.656	1.25
SE_{Avg}	28.604	0.003	0.007	0.003		0.002	0.004	0.006	0.059

Note: *df* = 114

The difference of χ^2 test did not reveal a statistically significant reduction of model misfit for either the unimputed data file (listwise deletion) or the imputed data files ($\Delta\chi^2_{Unimputed}[1] = 0.01$, p > 0.5; $\Delta\chi^2_{Imputed\ Avg}[1] = 2.923$, p = 0.087). I therefore, discarded this model and returned to Model A2.

The other MIs from Model A2 failed to reveal any theoretically sound alterations to the model. For example, Items 12 and 15, Items 1 and 4, measured different content areas and different underlying misconceptions. And although Items 13 and 14 both examined rational number content, they measured different underlying misconceptions. Therefore, I concluded that Model A2 was the best calibration of Model A possible with these data (Figure 25).

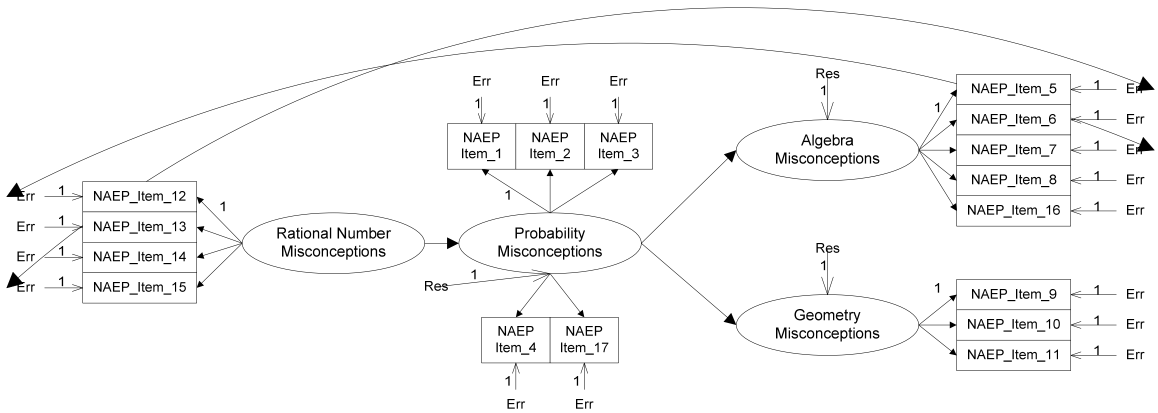


Figure 25. Final Structural Model A2

Validation of Model A2. To examine the convergent validity of Model A2, the goodness of fit statistics were computed based on the validation sample (Table 24).

Table 24
Model A2 Goodness of Fit Indices from Validation Samples

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 449)	621.32	0.98	0.86	0.025	[0.0098, 0.036]	0.081	0.65	0.741	1.50
1 (N = 556)	617.42	0.99	0.89	0.020	[0.000, 0.031]	0.070	0.67	0.745	1.14
2 (N = 580)	717.36	0.98	0.88	0.028	[0.018, 0.037]	0.076	0.66	0.735	1.27
3 (N = 575)	764.96	0.98	0.88	0.030	[0.020, 0.038]	0.078	0.66	0.735	1.30
4 (N = 567)	651.59	0.99	0.90	0.022	[0.0073, 0.032]	0.069	0.67	0.737	1.13
5 (N = 558)	701.17	0.98	0.88	0.028	[0.017, 0.037]	0.0742	0.66	0.735	1.26
Wtd. Avg.	689.92	0.98	0.89	0.026	[0.018, 0.033]	0.073	0.66	0.737	1.22
SE_{Avg}	28.762	0.003	0.004	0.002		0.002	0.003	0.002	0.040

Note: *df* = 115

The goodness of fit statistics were then compared to those based on the calibration sample using a *t*-test to compare the difference in the point estimates. Even though the population distribution of these fit indices are not always normally distributed, because of the central limit theorem, the sampling distribution around the point estimate will always be normally distributed, so a two-way, two-sample *t*-test is appropriate for comparisons of the imputed data sets (Table 25).

Table 25
Model A2 Comparison of Calibration and Validation Sample Fit Indices

Imputation	χ^2	CFI	GFI	RMSEA	SRMR	PGFI	PCFI	ECVI
Imputed Data Set *t* value	0.40	0.00	-0.88	-0.39	-0.50	0.00	1.57	0.82

No statistic from the imputed data sets was significantly different for the calibration and validation samples (i.e., all t values less than 1.96), indicating that Model A2 fit the validation and calibration samples equally well. I, therefore, concluded that the model had good convergent validity across samples.

Analysis of Model B

Calibration. Model B reversed the relationship specified in Model A for probability and rational number. In this model, misconceptions in rational number content filtered the influence of probability on the development of misconceptions in algebra and geometry. The goodness of fit indices suggested that this model fit the data well (Table 26).

Table 26
Model B Goodness of Fit Indices from Calibration Samples

Imputation	χ^2_a	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	607.05	0.99	0.87	0.023	[0.008, 0.034]	0.075	0.66	0.75	1.41
1 (N = 577)	756.98	0.97	0.87	0.032	[0.023, 0.040]	0.075	0.66	0.74	1.39
2 (N = 553)	682.40	0.99	0.89	0.018	[0.000, 0.029]	0.066	0.68	0.76	1.14
3 (N = 558)	744.63	0.98	0.87	0.031	[0.021, 0.040]	0.074	0.66	0.74	1.41
4 (N = 566)	649.92	0.98	0.88	0.026	[0.015, 0.035]	0.074	0.67	0.75	1.27
5 (N = 575)	710.91	0.99	0.88	0.024	[0.013, 0.034]	0.071	0.67	0.75	1.25
Wtd. Avg.	709.18	0.98	0.88	0.026	[0.017, 0.036]	0.072	0.67	0.748	1.29
SE_{Avg}	22.042	0.004	0.004	0.003		0.002	0.004	0.004	0.055

Note: *df* = 116

The unimputed data set called for the addition to Model B of the regression path from geometry to algebra (Table 27). This potential structural alteration seemed to offer the most substantive change to the hypothesized model. Therefore, this parameter was added to the model.

Table 27
Model B Maximum Modification Indices from Calibration Samples

Imputation	Maximum Modification Index	Associated Parameter to Add
0	165.58	BE (2, 3); Regression path from Geometry to Algebra
1	62.53	TD (4, 1); Covariance between Item 1 and 4 errors
2	44.08	TE (12, 7); Covariance between Item 7 and 11 errors
3	267.85	LY(5, 1); Crossloading, Rational Number to Item 5
4	46.16	TE(9, 4); Covariance between Item 15 and 16 errors
5	71.34	TE(5, 4); Covariance between Item 5 and 15 errors

Although the addition of this parameter to Model A resulted in an unstable model, it did not have such an effect on Model B.Even though the parameter represented the same regression path, its meaning within the model was quite different from that of Model A. In Model A, it represented dependence of algebra misconceptions on geometry misconceptions above and beyond its dependence on probability misconceptions with rational number misconceptions acting as the independent, exogenous variable. In Model B, this regression pathway represented the dependence of algebra misconceptions on geometry misconceptions above and beyond its dependence on rational number misconceptions while probability misconceptions acted as the independent, exogenous variable. Table 28 displays the resultant model fit indices.

Table 28

Model B2a Goodness of Fit Indices, Regress Algebra Misconceptions on Geometry Misconceptions

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	606.20	0.99	0.87	0.024	[0.0088, 0.035]	0.075	0.65	0.740	1.42
1 (N = 577)	756.28	0.97	0.87	0.032	[0.023, 0.041]	0.075	0.65	0.725	1.39
2 (N = 553)	676.54	0.99	0.90	0.017	[0.000, 0.028]	0.067	0.67	0.737	1.13
3 (N = 558)	738.19	0.98	0.87	0.031	[0.022, 0.040]	0.073	0.65	0.732	1.42
4 (N = 566)	644.92	0.98	0.88	0.026	[0.015, 0.035]	0.073	0.66	0.735	1.27
5 (N = 575)	703.87	0.99	0.88	0.024	[0.013, 0.034]	0.071	0.66	0.743	1.25
Wtd. Avg.	704.19	0.98	0.88	0.026	[0.016, 0.036]	0.072	0.66	0.734	1.29
SE_{Avg}	22.565	0.004	0.006	0.003		0.002	0.004	0.003	0.058

Note: *df* = 115

The difference of χ^2 test revealed a schism between the imputed data set averages and the unimputed data file. In the unimputed data file, the difference was not statistically significant ($\Delta\chi^2_{Unimputed}$ [1] = 0.85, p = 0.357). Using the weighted average of the five imputed data sets, the reduction of model misfit was significant ($\Delta\chi^2_{Imputed\ Avg}$[1] = 4.988, p = 0.026). To decide whether to retain the model, I considered that the MI that began this calibration came from the unimputed data set. So, I expected that any schism should have favored the unimputed data set rather than the imputed data sets. This reversal of effects suggested to me that the parameter did not affect the model the way the MI suggested it would and that the parameter may add instability to the model even though all the data sets converged. I therefore decided to reject this model change and returned to the original hypothesized Model B.

Other maximum MIs from the original Model B called for the addition of a crossloading from Rational Number misconceptions to Item 5 and error covariances between Items 1 and 4, 7 and 11, 5 and 15, and 15 and 16. I found no theoretical foundation for co-varying Items 1 and 4, Items 7 and 11, or Items 15 and 16. Items 5 and 15, on the other hand, both measured rational number meaning misconceptions; so, their covariance seemed plausible as well as the cross-loading between rational number misconceptions and Item 5. Since the cross-loading indicated the largest drop in model misfit, I chose to try it first. Table 29 displays the resultant fit indices.

Table 29
Model B2b Goodness of Fit Indices, Regress Item 5 on Rational Number Misconceptions

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	579.13	0.99	0.88	0.021	[0.000, 0.032]	0.075	0.66	0.743	1.36
1 (N = 577)	755.81	0.97	0.87	0.032	[0.023, 0.041]	0.075	0.66	0.736	1.38
2 (N = 553)	680.76	0.99	0.89	0.018	[0.000, 0.029]	0.066	0.67	0.745	1.14
3 (N = 558)	737.37	0.98	0.87	0.030	[0.020, 0.039]	0.074	0.66	0.743	1.38
4 (N = 566)	630.43	0.99	0.89	0.023	[0.010, 0.033]	0.073	0.67	0.745	1.21
5 (N = 575)	675.78	0.99	0.89	0.021	[0.0066, 0.031]	0.070	0.67	0.745	1.19
Wtd. Avg.	696.15	0.98	0.88	0.025	[0.015, 0.035]	0.072	0.67	0.743	1.26
SE_{Avg}	25.285	0.004	0.005	0.003		0.002	0.003	0.002	0.056

Note: *df* = 115

The difference of χ^2 test revealed a statistically significant reduction of model misfit for both the unimputed data file (listwise deletion) and the imputed data files ($\Delta\chi^2_{Unimputed}[1] = 27.92$, $p < 0.0001$; $\Delta\chi^2_{Imputed\ Avg}[1] = 13.029$, $p = 0.0003$). I therefore, retained Model B2b and examined its modification indices (Table 30).

Table 30
Model B2b Maximum Modification Indices

Imputation	Maximum Modification Index	Associated Parameter to Add
0	37.03	TD (3, 1); Covariance between Item 1 and 3 errors
1	62.45	TD (4, 1); Covariance between Item 1 and 4 errors
2	44.22	TE (12, 7); Covariance between Item 7 and 11 errors
3	125.79	TE (12, 6); Covariance between Item 6 and 11 errors
4	56.68	TE (12, 10); Covariance between Item 11 and 9 errors
5	40.62	TE (11, 10); Covariance between Item 9 and 10 errors

Items 1 and 4, Items 6 and 11, and Items 7 and 11 measured different content areas and different underlying misconceptions, so these three pairs of error covariances were excluded from consideration. Items 1 and 3 measured probability content knowledge but did not measure the same underlying misconception — Item 1 examined absolute/relative comparison misconceptions while Item 3 examined meaning of rational number misconceptions. Likewise, Items 9 and 11 both measured geometry knowledge, but Item 9 measured spatial reasoning misconceptions while Item 11 measured meaning of variable misconceptions. Therefore, these error covariances were considered theoretically marginal for inclusion in the model. Items 9 and 10, however, both measured geometry content knowledge and spatial reasoning misconceptions, so their error covariance was explored for possible inclusion in the model (Table 31).

Table 31
Model B3 Goodness of Fit Indices, Covary Item 9 and 10 Errors

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	575.78	0.99	0.88	0.021	[0.000, 0.033]	0.074	0.65	0.731	1.35
1 (N = 577)	724.69	0.98	0.88	0.030	[0.020, 0.039]	0.074	0.65	0.724	1.32
2 (N = 553)	676.54	0.99	0.89	0.019	[0.000, 0.030]	0.066	0.67	0.745	1.14
3 (N = 558)	712.22	0.98	0.88	0.029	[0.019, 0.038]	0.072	0.65	0.724	1.35
4 (N = 566)	622.22	0.99	0.89	0.023	[0.0095, 0.033]	0.072	0.66	0.734	1.19
5 (N = 575)	650.53	0.99	0.89	0.020	[0.000, 0.030]	0.068	0.67	0.745	1.16
Wtd. Avg.	677.24	0.99	0.89	0.024	[0.008, 0.016]	0.070	0.66	0.734	1.23
SE_{Avg}	21.237	0.003	0.003	0.003		0.002	0.005	0.005	0.048

Note: *df* = 114

The difference of χ^2 test revealed a statistically significant decrease in model misfit in the imputed data sets ($\Delta\chi^2_{\text{Imputed Avg}}[1] = 18.907$, $p < 0.0001$). The reduction in model misfit, however, was significant only within a 93% confidence interval

($\Delta\chi^2_{Unimputed}$ [1] = 3.35, p = 0.067). Since the modification reduced a significant amount of misfit across the imputed data sets and nearly significant in the unimputed data sets, I concluded that the modification should be retained. Therefore, the modification indices for Model B3 were examined for further calibration (Table 32).

Table 32
Model B3 Maximum Modification Indices

Imputation	Maximum Modification Index	Associated Parameter to Add
0	38.01	TD (3, 1); Covariance between Item 1 and 3 errors
1	62.40	TD (4, 1); Covariance between Item 1 and 4 errors
2	44.31	TE (12, 7); Covariance between Item 7 and 11 errors
3	78.97	TE (11, 7); Covariance between Item 7 and 10 errors
4	38.25	TE (3, 2); Covariance between Item 13 and 14 errors
5	39.06	TH (5, 8); Covariance between Item 8 and 17 errors

Items 1 and 3, Items 1 and 4, and Items 13 and 14 measured similar content knowledge but different underlying misconceptions, so these error covariances were considered theoretically marginal for inclusion in the model. Items 7 and 10, Items 7 and 11, and Items 8 and 17 measured different content knowledge and different underlying misconceptions, so these error covariances were considered theoretically poor for inclusion in the model. Therefore, I concluded that Model B3 (Figure 26) represented the best calibration of the hypothesized model with these data.

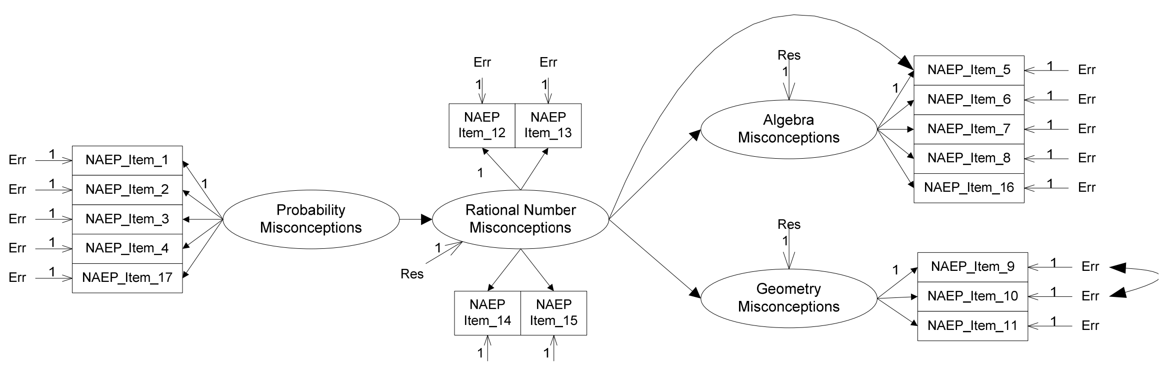


Figure 26. Final Structural Model B3.

Validation of Model B3. To examine the convergent validity of Model B3, the goodness of fit statistics were computed based on the validation sample (Table 33).

Table 33
Model B3 Goodness of Fit Indices from Validation Samples

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 449)	628.51	0.98	0.86	0.026	[0.011, 0.037]	0.081	0.64	0.729	1.51
1 (N = 556)	592.18	0.99	0.90	0.018	[0.000, 0.029]	0.068	0.67	0.737	1.12
2 (N = 580)	704.41	0.98	0.88	0.028	[0.017, 0.037]	0.074	0.66	0.735	1.25
3 (N = 575)	755.89	0.98	0.88	0.029	[0.019, 0.038]	0.077	0.66	0.735	1.28
4 (N = 567)	612.77	0.99	0.90	0.019	[0.000, 0.030]	0.067	0.67	0.737	1.09
5 (N = 558)	665.66	0.98	0.89	0.026	[0.015, 0.035]	0.072	0.66	0.727	1.22
Wtd. Avg.	665.46	0.98	0.89	0.024	[0.015, 0.032]	0.072	0.66	0.734	1.19
SE_{Avg}	33.392	0.003	0.005	0.003		0.002	0.003	0.002	0.041

Note: *df* = 114

The goodness of fit statistics were then compared to those based on the calibration sample using a *t*-test to compare the difference in the point estimates (Table 34).

Table 34
Model B3 Comparison of Calibration and Validation Sample Fit Indices

Imputation	χ^2	CFI	GFI	RMSEA	SRMR	PGFI	PCFI	ECVI
Imputed *t* value	0.42	0.67	-0.97	0.00	-1.00	-0.97	0.00	0.92

No statistic from the imputed data sets was significantly different for the calibration and validation samples (i.e., all t values less than 1.96), indicating that Model B3 fit the validation and calibration samples equally well. I, therefore, concluded that the model had good convergent validity across samples.

Analysis of Model C

Model C specified misconceptions in rational number and probability content areas as covarying independent variables, each directly influencing the development of misconceptions in algebra and geometry content areas.

Finding a structurally stable Model C. Imputations 1, 2, and 4 for Model C would not converge with the hypothesized model. Using intermediate reported values as new starting points, the models were run several times, trying to reach convergence. Instead of converging, the models continued to diverge (Table 35).

Table 35
Model C Goodness of Fit Indices from Calibration Samples

Imputation	χ^2_a	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	603.47	0.99	0.87	0.024	[0.009, 0.035]	0.075	0.65	0.74	1.42
1 (N = 577)				*Hypothesized Model Would not Converge*					
2 (N = 553)				*Hypothesized Model Would not Converge*					
3 (N = 558)	722.31	0.98	0.87	0.030	[0.020, 0.039]	0.073	0.65	0.73	1.37
4 (N = 566)				*Hypothesized Model Would not Converge*					
5 (N = 575)	710.29	0.99	0.88	0.025	[0.013, 0.034]	0.071	0.66	0.74	1.26
Wtd. Avg.	716.21	0.99	0.88	0.027	[0.021, 0.033]	0.072	0.66	0.74	1.31
SE_{Avg}	8.500	0.007	0.007	0.004		0.001	0.007	0.007	0.078

Note: *df* = 114

An examination of the intermediate values revealed that the parameter estimates for the regression weights from probability to geometry, from rational number to algebra, and from probability to algebra differed greatly from all other estimates. Therefore, modifications were made to the model, starting with removing the regression pathway from probability to geometry. With this change, all imputations except Imputation 4 converged (Table 36).

Table 36
Model C2a Calibration Samples Goodness of Fit Indices

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	603.71	0.99	0.87	0.024	[0.0084, 0.035]	0.075	0.66	0.751	1.41
1 (N = 577)	754.77	0.97	0.87	0.033	[0.024, 0.041]	0.075	0.65	0.725	1.40
2 (N = 553)	679.23	0.99	0.89	0.018	[0.000, 0.029]	0.066	0.67	0.745	1.15
3 (N = 558)	723.33	0.98	0.87	0.029	[0.019, 0.038]	0.073	0.66	0.743	1.36
4 (N = 566)					*Model C2a did not Converge*				
5 (N = 575)	710.32	0.99	0.88	0.025	[0.013, 0.034]	0.071	0.66	0.743	1.26
Wtd. Avg.	717.26	0.98	0.88	0.026	[0.021, 0.031]	0.071	0.66	0.739	1.29
SE_{Avg}	9.220	0.008	0.008	0.003		0.002	0.000	0.006	0.074

Note: *df* = 115

Since Imputation 4 did not converge, the original hypothesized Model C was adjusted again by removing the regression pathway from rational number to algebra (Table 37).

Table 37

Model C2b Calibration Samples Goodness of Fit Indices

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	606.39	0.99	0.87	0.024	[0.0098, 0.035]	0.075	0.65	0.74	1.42
1 (N = 577)				*Model C2b did not Converge*					
2 (N = 553)	678.18	0.99	0.89	0.019	[0.000, 0.030]	0.066	0.67	0.75	1.15
3 (N = 558)	762.34	0.98	0.87	0.031	[0.022, 0.040]	0.075	0.65	0.73	1.41
4 (N = 566)	645.32	0.98	0.88	0.025	[0.014, 0.035]	0.074	0.66	0.74	1.25
5 (N = 575)				*Model C2b did not Converge*					
Wtd. Avg.	695.09	0.98	0.88	0.025	[0.018, 0.032]	0.072	0.66	0.737	1.27
SE_{Avg}	30.178	0.003	0.005	0.003		0.002	0.005	0.003	0.066

Note: *df* = 115

Since Imputations 1 and 5 did not converge, the original hypothesized Model C (Model C3a) was adjusted again by removing both regression pathways, rational number to algebra and probability to geometry (Table 38).

Table 38

Model C3a Fit Indices, Remove Rational Number to Algebra and Probability to Geometry Regression Paths

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	607.07	0.99	0.87	0.024	[0.0086, 0.035]	0.075	0.66	0.751	1.42
1 (N = 577)	759.38	0.97	0.87	0.032	[0.023, 0.041]	0.075	0.66	0.736	1.40
2 (N = 553)	679.80	0.99	0.89	0.018	[0.000, 0.029]	0.066	0.68	0.756	1.15
3 (N = 558)	762.49	0.98	0.87	0.031	[0.021, 0.040]	0.075	0.66	0.743	1.40
4 (N = 566)	648.89	0.98	0.88	0.025	[0.014, 0.035]	0.074	0.67	0.746	1.26
5 (N = 575)	727.47	0.99	0.88	0.025	[0.014, 0.035]	0.072	0.67	0.754	1.27
Wtd. Avg.	715.85	0.98	0.88	0.026	[0.017, 0.036]	0.072	0.67	0.747	1.30
SE_{Avg}	24.986	0.004	0.004	0.003		0.002	0.004	0.004	0.053

Note: *df* = 116

The difference of χ^2 test did not reveal statistically significant reduction of model misfit for either the unimputed data file (listwise deletion) or the imputed data files ($\Delta\chi^2_{Unimputed}[2] = 1.61$, p = 0.477; $\Delta\chi^2_{Imputed\ Avg}[2] = 0.89$, p > 0.5). Alternatively, the other set of regression pathways was removed, rational number to geometry and probability to algebra (Model C3b; Table 39).

Table 39

Model C3b Fit Indices, Remove Rational Number to Geometry and Probability to Algebra Regression Paths

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	608.68	0.99	0.87	0.024	[0.0085, 0.035]	0.075	0.66	0.75	1.42
1 (N = 577)	754.60	0.97	0.87	0.032	[0.023, 0.040]	0.076	0.66	0.74	1.39
2 (N = 553)	679.95	0.99	0.89	0.017	[0.000, 0.029]	0.066	0.68	0.76	1.14
3 (N = 558)	771.63	0.98	0.87	0.032	[0.022, 0.040]	0.075	0.66	0.74	1.43
4 (N = 566)	652.76	0.98	0.88	0.026	[0.015, 0.035]	0.074	0.67	0.75	1.28
5 (N = 575)	723.82	0.99	0.88	0.025	[0.014, 0.034]	0.072	0.67	0.75	1.27
Wtd. Avg.	716.74	0.98	0.88	0.026	[0.016, 0.037]	0.073	0.67	0.747	1.30
SE_{Avg}	24.929	0.004	0.004	0.003		0.002	0.004	0.004	0.057

Note: *df* = 116

The difference of χ^2 test did not reveal statistically significant reduction of model misfit for either the unimputed data file (listwise deletion) or the imputed data files ($\Delta\chi^2_{Unimputed}$ [2] = 3.6, p = 0.165; $\Delta\chi^2_{Imputed\ Avg}$[2] = 5.21, p = 0.074). However, both models added stability across imputations, so they were deemed superior to the original hypothesized model. Furthermore, neither model provided a statistically better fit, so both were retained for the synthesizing of the structural model calibration results.

Calibration of Model C3. Model C specified Rational Number and Probability Misconceptions as covarying independent variables. Model C3A specified the regression of Algebra on Probability Misconceptions and of Geometry on Rational Number Misconceptions. Model C3B specified the regression of Geometry on Probability Misconceptions and of Algebra on Rational Number Misconceptions. Since both models emerged from the original hypothesized model and neither model contained a significantly lower amount of model misfit, both models were calibrated using their respective modification indices (Table 40).

Table 40
Model C3 Maximum Modification Indices from Calibration Samples

Model, Imputation	Maximum Modification Index	Associated Parameter to Add
Model C3-A, 0	89.94	LY(3, 2); Crossloading, Geometry to Item 7
1	53.42	TD (4, 1); Covariance between Item 1 and 4 errors
2	45.28	TE(8, 3); Covariance between Item 7 and 11 errors
3	106.69	TE(8, 2); Covariance between Item 6 and 11 errors
4	52.39	TE(8, 6); Covariance between Item 9 and 11 errors
5	64.77	TH(9, 1); Covariance between Item 5 and 15 errors
Model C3-B, 0	31.54	TH(1, 7); Covariance between Item 5 and 15 errors
1	58.32	TD(4, 1); Covariance between Item 1 and 4 errors
2	343.10	TH(6, 8); Covariance between Item 11 and 12 errors
3	61.68	TD(4, 1); Covariance between Item 1 and 4 errors
4	45.46	TH(9, 5); Covariance between Item 15 and 16 errors
5	91.97	TH(9, 1); Covariance between Item 5 and 15 errors

Calibration of Model C3A. No theoretical foundation supported the crossloading from Geometry to Item 7, the covariance between Item 1 and 4 errors, Item 7 and 11 errors, or Item 6 and 11 errors. The covariance between Item 9 and 11 errors was considered theoretically marginal since those items measured the same type of content knowledge but different underlying misconceptions. Likewise, Items 5 and 15 measured the same underlying misconception (rational number meaning) but different content knowledge. I chose to try the covariance between Item 5 and 15 errors because the alignment of underlying misconceptions seemed more consistent with the present study purpose. Table 41 displays the goodness of fit statistics for the resultant model.

Table 41
Model C3Aii Goodness of Fit Indices, Covary Item 5 and 15 Errors

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	579.28	0.99	0.88	0.021	[0.000, 0.033]	0.075	0.66	0.743	1.36
1 (N = 577)	757.67	0.97	0.87	0.032	[0.024, 0.041]	0.075	0.65	0.725	1.39
2 (N = 553)	678.58	0.99	0.89	0.019	[0.000, 0.030]	0.066	0.67	0.745	1.15
3 (N = 558)	755.37	0.98	0.87	0.030	[0.020, 0.039]	0.075	0.66	0.743	1.37
4 (N = 566)	629.57	0.99	0.89	0.023	[0.0095, 0.032]	0.073	0.67	0.745	1.20
5 (N = 575)	687.99	0.99	0.89	0.022	[0.0087, 0.032]	0.070	0.67	0.745	1.20
Wtd. Avg.	701.97	0.98	0.88	0.025	[0.016, 0.034]	0.072	0.66	0.741	1.26
SE_{Avg}	27.316	0.004	0.005	0.003		0.002	0.004	0.004	0.055

Note: *df* = 115

The difference of χ^2 test revealed a statistically significant reduction of model misfit for both the unimputed data file (listwise deletion) and the imputed data files ($\Delta\chi^2$

$_{\text{Unimputed}}[1] = 27.79$, $p < 0.0001$; $\Delta\chi^2_{\text{Imputed Avg}}[1] = 13.881$, $p = 0.0002$). I therefore, retained Model C3Aii and examined its modification indices (Table 42).

Table 42
Model C3Aii Maximum Modification Indices

Imputation	Maximum Modification Index	Associated Parameter to Add
0	76.22	LY (3, 2); Crossloading between Geometry and Item 7
1	53.02	TD (4, 1); Covariance between Item 1 and 4 errors
2	45.29	TE (8, 3); Covariance between Item 7 and 11 errors
3	105.95	TE (8, 2); Covariance between Item 6 and 11 errors
4	375.57	TH (6, 3); Covariance between Item 7 and 12 errors
5	40.75	TE (7, 6); Covariance between Item 9 and 10 errors

The potential modifications of crossloading Geometry to Item 7, covarying Items 1 and 4, Items 7 and 11, and Items 6 and 11 were considered theoretically weak because there was no shared content area or underlying modifications. Items 7 and 12 measured the same underlying misconception in different content areas, so I considered the addition of their covariance to be theoretically marginal. Items 9 and 10, on the other hand, measured the same content area and underlying misconception, so I chose to add their covariance to the model (Table 43).

Table 43
Model C3Aiii Goodness of Fit Indices, Covary Item 9 and 10 Errors

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	575.55	0.99	0.88	0.021	[0.000, 0.033]	0.074	0.65	0.731	1.36
1 (N = 577)	726.56	0.98	0.88	0.030	[0.021, 0.039]	0.074	0.65	0.724	1.32
2 (N = 553)	673.93	0.99	0.89	0.019	[0.000, 0.030]	0.066	0.67	0.745	1.15
3 (N = 558)	729.14	0.98	0.88	0.029	[0.019, 0.038]	0.073	0.65	0.724	1.34
4 (N = 566)	621.79	0.99	0.89	0.022	[0.009, 0.032]	0.072	0.66	0.734	1.19
5 (N = 575)	662.61	0.99	0.89	0.021	[0.005, 0.031]	0.069	0.66	0.734	1.17
Wtd. Avg.	682.82	0.99	0.89	0.024	[0.016, 0.032]	0.071	0.66	0.732	1.23
SE_{Avg}	22.736	0.003	0.003	0.002		0.002	0.004	0.004	0.045

Note: *df* = 114

The difference of χ^2 test revealed a statistically significant decrease in model misfit in the imputed data sets ($\Delta\chi^2_{\text{Imputed Avg}}[1] = 19.143$, $p < 0.0001$). The reduction in model misfit, however, was significant only within a 94% confidence interval for the unimputed data set ($\Delta\chi^2_{\text{Unimputed}}[1] = 3.73$, $p = 0.053$). Since the modification reduced a significant amount of misfit across the imputed data sets and nearly significant in the unimputed data sets, I concluded that the modification should be retained. Therefore, the modification indices for Model C3Aiii were examined for further calibration (Table 44).

Table 44
Model C3Aiii Maximum Modification Indices

Imputation	Maximum Modification Index	Associated Parameter to Add
0	32.26	TD (8, 6); Covariance between Item 12 and 14 errors
1	53.13	TD (4, 1); Covariance between Item 1 and 4 errors
2	45.28	TE (8, 3); Covariance between Item 7 and 11 errors
3	65.74	TE (7, 3); Covariance between Item 7 and 10 errors
4	39.93	TD (8, 7); Covariance between Item 13 and 14 errors
5	39.26	TH (5, 4); Covariance between Item 8 and 17 errors

Of the potential modifications for Model C3Aiii, none were theoretically relevant except the covariance of Item 12 and 14 errors and of Item 13 and 14 errors. Both of these item pairs measured the same content area but different underlying misconceptions. Therefore, I considered these modifications to be theoretically marginal. Furthermore, the magnitudes of MIs appeared small relative to previous models, so I concluded that Model C3Aiii represented the best calibration of Model C3A for these data (Figure 27).

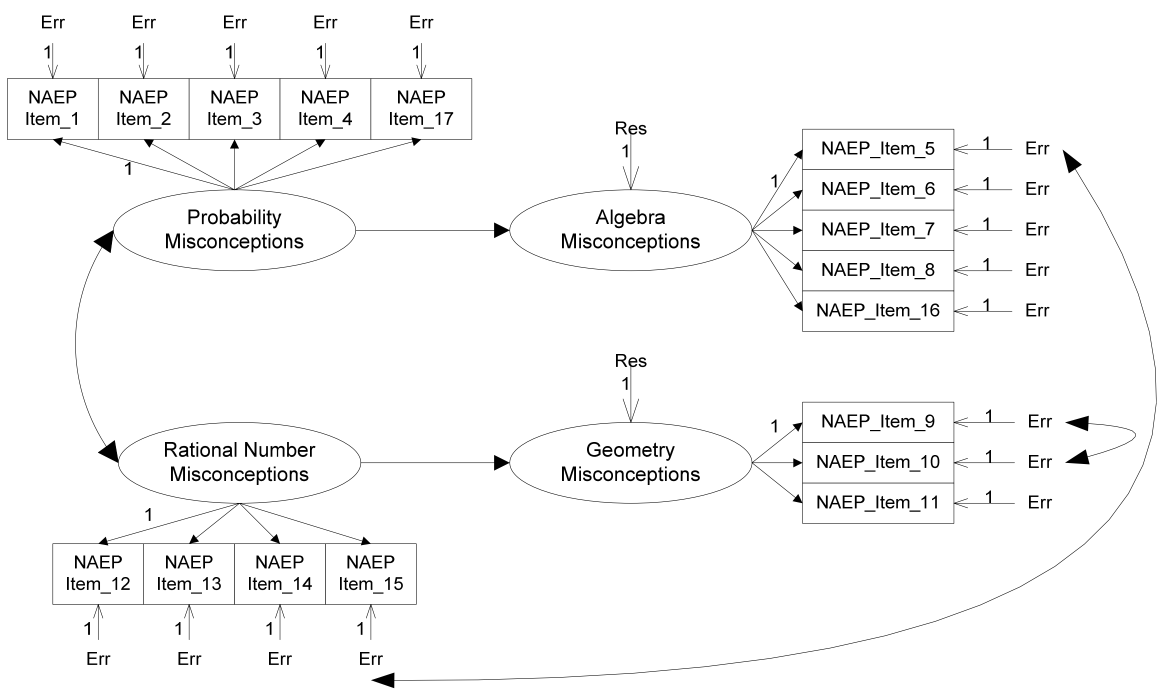


Figure 27. Final Structural Model C3Aiii.

Validation of Model C3Aiii. To examine the convergent validity of Model C3Aiii, the goodness of fit statistics were computed based on the validation sample (Table 45).

Table 45
Model C3Aiii Goodness of Fit Indices from Validation Samples

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 449)	626.37	0.98	0.86	0.026	[0.012, 0.037]	0.081	0.64	0.729	1.52
1 (N = 556)	615.57	0.99	0.89	0.020	[0.000, 0.030]	0.070	0.67	0.745	1.14
2 (N = 580)	727.21	0.98	0.88	0.029	[0.019, 0.037]	0.076	0.66	0.735	1.27
3 (N = 575)	753.28	0.98	0.88	0.029	[0.019, 0.038]	0.077	0.66	0.735	1.28
4 (N = 567)	639.74	0.99	0.90	0.021	[0.0064, 0.031]	0.068	0.67	0.737	1.12
5 (N = 558)	687.05	0.98	0.88	0.027	[0.017, 0.037]	0.073	0.66	0.735	1.25
Wtd. Avg.	683.92	0.98	0.89	0.025	[0.018, 0.032]	0.073	0.664	0.737	1.21
SE_{Avg}	28.851	0.003	0.004	0.002		0.002	0.003	0.002	0.038

Note: *df* = 114

The goodness of fit statistics were then compared to those based on the calibration sample using a *t*-test to compare the difference in the point estimates (Table 46).

Table 46
Model C3Aiii Comparison of Calibration and Validation Sample Fit Indices

Imputation	χ^2	CFI	GFI	RMSEA	SRMR	PGFI	PCFI	ECVI
Imputed Data Set *t* value	-0.04	0.67	0.00	-0.50	-1.00	-1.70	-1.58	0.55

No statistic from the imputed data sets was significantly different for the calibration and validation samples (i.e., all t values less than 1.96), indicating that Model C3Aiii fit the validation and calibration samples equally well. I, therefore, concluded that the model had good convergent validity across samples.

Calibration of Model C3B. The potential modifications of covarying Item 11 and 12 errors, Item 1 and 4 errors, and Items 15 and 16 errors were considered theoretically weak because they did not measure the same content knowledge or underlying misconception. Because Items 5 and 15 measured the same underlying misconception, the covariance of their errors was added to the model (Table 47).

Table 47
Model C3Bii Goodness of Fit Indices, Covary Item 5 and 15 Errors

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	580.28	0.99	0.88	0.021	[0.000, 0.033]	0.075	0.66	0.743	1.36
1 (N = 577)	752.43	0.97	0.87	0.032	[0.023, 0.041]	0.075	0.66	0.736	1.31
2 (N = 553)	678.98	0.99	0.89	0.018	[0.000, 0.029]	0.066	0.67	0.745	1.14
3 (N = 558)	764.96	0.98	0.87	0.031	[0.021, 0.040]	0.075	0.65	0.732	1.40
4 (N = 566)	630.49	0.99	0.89	0.023	[0.0099, 0.033]	0.073	0.67	0.745	1.20
5 (N = 575)	687.72	0.99	0.89	0.022	[0.0083, 0.032]	0.070	0.67	0.745	1.20
Wtd. Avg.	703.00	0.98	0.88	0.025	[0.015, 0.035]	0.072	0.66	0.741	1.25
SE_{Avg}	27.783	0.004	0.005	0.003		0.002	0.004	0.003	0.052

Note: *df* = 115

The difference of χ^2 test revealed a statistically significant decrease in model misfit for both the unimputed data set (listwise deletion) and the imputed data sets ($\Delta\chi^2_{Unimputed}[1] = 28.4$, $p < 0.0001$; $\Delta\chi^2_{Imputed\ Avg}[1] = 13.741$, $p = 0.0002$). Therefore, Model C3Bii was retained, and the modification indices were examined (Table 48).

Table 48
Model C3Bii Maximum Modification Indices

Imputation	Maximum Modification Index	Associated Parameter to Add
0	32.22	TH (1, 7); Covariance between Item 1 and 10 errors
1	58.08	TD (4, 1); Covariance between Item 1 and 4 errors
2	127.66	TH (6, 8); Covariance between Item 11 and 12 errors
3	61.75	TD (4, 1); Covariance between Item 1 and 4 errors
4	37.65	TD (8, 7); Covariance between Item 13 and 14 errors
5	52.48	TH (8, 1); Covariance between Item 5 and 14 errors

Items 1 and 10, Items 1 and 4, Items 5 and 14, and Items 11 and 12 measured different content knowledge and underlying misconceptions, so they were discarded as potential modifications. Items 13 and 14, on the other hand, both measured rational number content while examining different underlying misconceptions. In this case, the underlying misconceptions were rational number meaning and absolute/relative comparisons, and both examined rational number content. The relationship between these two misconceptions, therefore, warranted the addition of this covariance parameter to the model (Table 49).

Table 49
Model C3Biii Goodness of Fit Indices, Covary Item 13 and 14 Errors

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	579.49	0.99	0.88	0.022	[0.000, 0.033]	0.075	0.65	0.73	1.36
1 (N = 577)	751.81	0.97	0.87	0.032	[0.024, 0.041]	0.075	0.65	0.72	1.38
2 (N = 553)	678.80	0.99	0.89	0.019	[0.000, 0.030]	0.066	0.67	0.75	1.14
3 (N = 558)	764.66	0.98	0.87	0.031	[0.022, 0.040]	0.075	0.65	0.73	1.40
4 (N = 566)	592.02	0.99	0.89	0.019	[0.000, 0.030]	0.071	0.67	0.75	1.14
5 (N = 575)	686.72	0.99	0.89	0.022	[0.0093, 0.032]	0.070	0.66	0.73	1.21
Wtd. Avg.	694.87	0.98	0.88	0.025	[0.014, 0.035]	0.071	0.66	0.736	1.25
SE_{Avg}	34.474	0.004	0.005	0.003		0.002	0.005	0.007	0.064

Note: *df* = 114

The difference of χ^2 test revealed a statistically significant decrease in model misfit in the imputed data sets ($\Delta\chi^2_{Imputed\ Avg}[1] = 8.121$, p = 0.004). The reduction in model misfit, however, was non-significant for the unimputed data ($\Delta\chi^2_{Unimputed}[1] = 0.79$, p = 0.374). Furthermore, the ECVI values increased, suggesting that whatever model misfit was eliminated by the new parameter was the result of overfitting the model to a sample. Therefore, I removed this parameter and retained Model C3Bii as the best calibration of this model for these data (Figure 28).

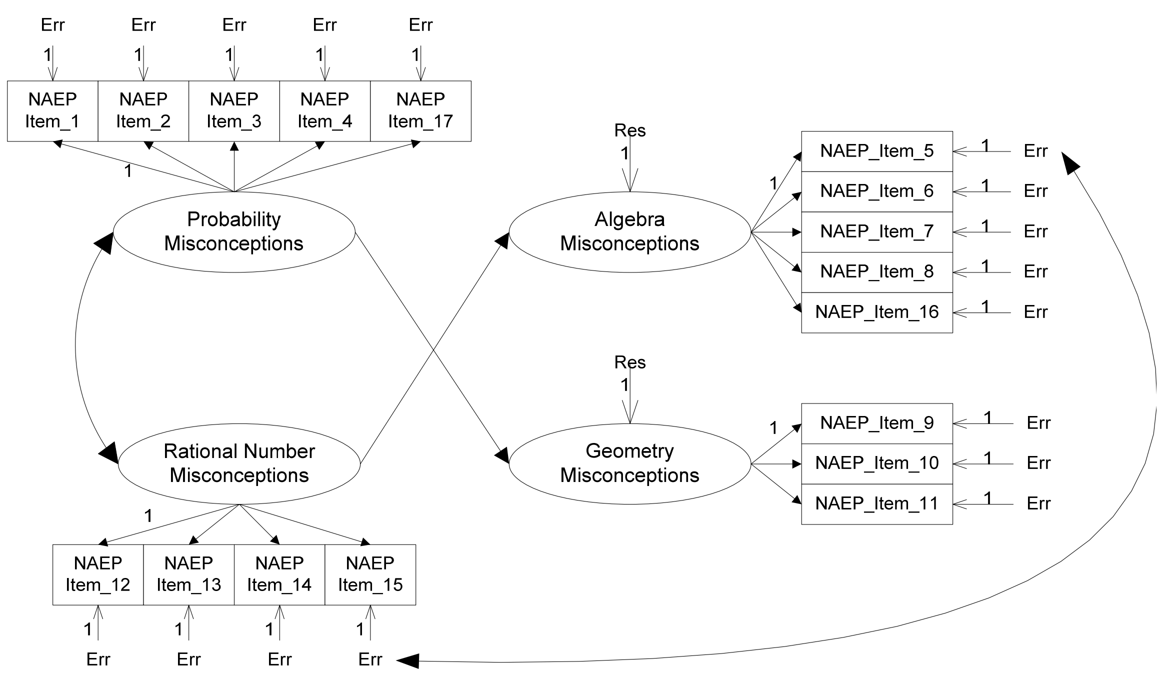


Figure 28. Final Structural Model C3Bii.

Validation of Model C3Bii. To examine the convergent validity of Model C3Aiii, the goodness of fit statistics were computed based on the validation sample (Table 50).

Table 50
Model C3Bii Goodness of Fit Indices from Validation Samples

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 449)	627.99	0.98	0.86	0.025	[0.0093, 0.036]	0.081	0.65	0.741	1.50
1 (N = 556)	611.02	0.99	0.89	0.019	[0.000, 0.030]	0.070	0.67	0.745	1.13
2 (N = 580)	754.73	0.98	0.88	0.029	[0.020, 0.038]	0.077	0.66	0.735	1.30
3 (N = 575)	764.78	0.98	0.88	0.029	[0.020, 0.038]	0.078	0.66	0.735	1.30
4 (N = 567)	641.47	0.99	0.90	0.020	[0.0037, 0.031]	0.068	0.67	0.737	1.11
5 (N = 558)	698.49	0.98	0.88	0.027	[0.017, 0.037]	0.074	0.66	0.735	1.26
Wtd. Avg.	693.31	0.98	0.89	0.025	[0.017, 0.033]	0.073	0.66	0.737	1.22
SE_{Avg}	33.881	0.003	0.004	0.002		0.002	0.003	0.002	0.047

Note: *df* = 115

The goodness of fit statistics were then compared to those from the calibration sample using a *t*-test to compare the difference in the point estimates (Table 51).

Table 51
Model C3Bii Comparison of Calibration and Validation Sample Fit Indices

Imputation	χ^2	CFI	GFI	RMSEA	SRMR	PGFI	PCFI	ECVI
Imputed Data Set *t* value	0.31	0.00	-0.88	0.00	-0.50	0.00	1.57	0.63

No statistic from the imputed data sets was significantly different for the calibration and validation samples (i.e., all t values less than 1.96), indicating that Model C3Bii fit the validation and calibration samples equally well. I, therefore, concluded that the model had good convergent validity across samples.

Analysis of Model D

Calibration. The original hypothesized Model D specified Rational Number Misconceptions as the sole independent variable with Probability, Algebra, and Geometry Misconceptions acting as dependent variables. The goodness of fit indices indicated an excellent fit (Table 52).

Table 52
Model D Goodness of Fit Indices from Calibration Samples

Imputation	χ^2_a	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	607.05	0.99	0.87	0.023	[0.008, 0.034]	0.075	0.66	0.751	1.41
1 (N = 577)	756.98	0.97	0.87	0.032	[0.023, 0.040]	0.075	0.66	0.736	1.39
2 (N = 553)	682.40	0.99	0.89	0.018	[0.000, 0.029]	0.066	0.68	0.756	1.14
3 (N = 558)	744.63	0.98	0.87	0.031	[0.021, 0.040]	0.074	0.66	0.743	1.41
4 (N = 566)	649.92	0.98	0.88	0.026	[0.015, 0.035]	0.074	0.67	0.746	1.27
5 (N = 575)	710.91	0.99	0.88	0.024	[0.013, 0.034]	0.071	0.67	0.754	1.25
Wtd. Avg.	709.18	0.98	0.88	0.026	[0.017, 0.036]	0.072	0.67	0.747	1.29
SE_{Avg}	22.042	0.004	0.004	0.003		0.002	0.004	0.004	0.055

Note: *df* = 116

The maximum MIs for Model D called for the addition of several error covariance terms (Table 53).

Table 53
Model D Maximum Modification Indices from Calibration Samples

Imputation	Maximum Modification Index	Associated Parameter to Add
0	37.24	TE(3,1); Covariance between Item 1 and 3 errors
1	62.53	TE(4,1); Covariance between Item 1 and 4 errors
2	44.08	TE(13,8); Covariance between Item 7 and 11 errors
3	113.58	TE(13,7); Covariance between Item 6 and 11 errors
4	46.16	TH(4, 10); Covariance between Item 15 and 16 errors
5	71.34	TH(4, 6); Covariance between Item 5 and 15 errors

The covariances of the error terms for Items 7 and 11, Items 6 and 11, and Items 15 and 16 were considered theoretically weak because they shared neither common content knowledge nor underlying misconception. The covariances of the error terms for Items 1 and 3 and Items 1 and 4 were considered theoretically marginal because they measured the same content knowledge but not the same underlying misconception. The covariance of the error terms for Items 5 and 15 was also considered theoretically marginal because the items measured the same underlying misconception but not the same content knowledge. To decide whether to add a parameter, and if so, which one, I also considered that Item 3 measured Rational Number Meaning misconceptions while Item 1 measured Absolute/Relative Comparison misconceptions, two misconceptions that are closely related. Finally, I considered that the covariance of Items 5 and 15 had been used in previous models to reduce a statistically significant amount of model misfit, that the MI for Items 5 and 15 was the second highest across the data sets, and that the MI for Items 1 and 3 was the lowest across the data sets. Therefore, I chose to add the covariance between Item 5 and 15 error terms to Model D (Table 54).

Table 54
Model D2 Goodness of Fit Indices, Covary Item 5and 15 Errors

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	579.13	0.99	0.88	0.021	[0.000, 0.032]	0.075	0.66	0.743	1.36
1 (N = 577)	755.81	0.97	0.87	0.032	[0.023, 0.041]	0.075	0.66	0.736	1.38
2 (N = 553)	680.76	0.99	0.89	0.018	[0.000, 0.029]	0.066	0.67	0.745	1.14
3 (N = 558)	737.37	0.98	0.87	0.030	[0.020, 0.039]	0.074	0.66	0.743	1.38
4 (N = 566)	630.43	0.99	0.89	0.023	[0.010, 0.033]	0.073	0.67	0.745	1.21
5 (N = 575)	675.78	0.99	0.89	0.021	[0.0066, 0.031]	0.070	0.67	0.745	1.19
Wtd. Avg.	696.15	0.98	0.88	0.025	[0.015, 0.035]	0.072	0.67	0.743	1.26
SE_{Avg}	25.285	0.004	0.005	0.003		0.002	0.003	0.002	0.056

Note: *df* = 115

The difference of χ^2 test revealed a statistically significant decrease in model misfit for both the unimputed data set (listwise deletion) and the imputed data sets ($\Delta\chi^2_{Unimputed}[1] = 27.92$, $p < 0.0001$; $\Delta\chi^2_{Imputed\ Avg}[1] = 13.031$, $p = 0.0003$). Therefore, Model D2 was retained, and the modification indices were examined (Table 55).

Table 55
Model D2 Maximum Modification Indices from Calibration Samples

Imputation	Maximum Modification Index	Associated Parameter to Add
0	37.03	TE(3,1); Covariance between Item 1 and 3 errors
1	62.45	TE(4,1); Covariance between Item 1 and 4 errors
2	44.22	TE(13,8); Covariance between Item 7 and 11 errors
3	125.79	TE(13,7); Covariance between Item 6 and 11 errors
4	56.68	TE(13, 11); Covariance between Item 9 and 11 errors
5	40.62	TE(12, 11); Covariance between Item 9 and 10 errors

The covariance of Item 9 and 10 error terms offered the theoretically strongest adjustment to Model D2. Table 56 displays the goodness of fit statistics for the new model resulting from this parameter (Model D3).

Table 56
Model D3 Goodness of Fit Indices, Covary Item 9 and 10 Errors

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	575.78	0.99	0.88	0.021	[0.000, 0.033]	0.074	0.65	0.731	1.35
1 (N = 577)	724.69	0.98	0.88	0.030	[0.020, 0.039]	0.074	0.65	0.724	1.32
2 (N = 553)	676.54	0.99	0.89	0.019	[0.000, 0.030]	0.066	0.67	0.745	1.14
3 (N = 558)	712.22	0.98	0.88	0.029	[0.019, 0.038]	0.072	0.65	0.724	1.35
4 (N = 566)	622.22	0.99	0.89	0.023	[0.0095, 0.033]	0.072	0.66	0.734	1.19
5 (N = 575)	650.53	0.99	0.89	0.020	[0.000, 0.030]	0.068	0.67	0.745	1.16
Wtd. Avg.	677.24	0.99	0.89	0.024	[0.016, 0.033]	0.070	0.66	0.734	1.23
SE_{Avg}	21.237	0.003	0.003	0.003		0.002	0.005	0.005	0.048

Note: $df = 114$

The difference of χ^2 test revealed a statistically significant decrease in model misfit in the imputed data sets ($\Delta\chi^2_{\text{Imputed Avg}}[1] = 18.907$, $p < 0.0001$). The reduction in model misfit, however, was significant only within a 93% confidence interval for the unimputed data set ($\Delta\chi^2_{\text{Unimputed}}[1] = 3.35$, $p = 0.067$). To decide whether to retain the model, I considered three characteristics of the analysis: (1) The large amount of model misfit removed across the imputed data sets, (2) The proximity of the significance level of the unimputed data set to a 95% confidence interval, and (3) The theoretical strength of the additional parameter. I concluded, therefore, that the modification should be retained and the modification indices for Model D3 were examined for further calibration (Table 57).

Table 57
Model D3 Maximum Modification Indices from Calibration Samples

Imputation	Maximum Modification Index	Associated Parameter to Add
0	2268.68	LY (1,3); Crossloading, Geometry to Item 1
1	62.40	TE(4,1); Covariance between Item 1 and 4 errors
2	44.31	TE(13,8); Covariance between Item 7 and 11 errors
3	78.97	TE(12,8); Covariance between Item 7 and 10 errors
4	38.25	TD(3, 2); Covariance between Item 13 and 14 errors
5	39.06	TE(9, 5); Covariance between Item 8 and 17 errors

I found no theoretical support for adding the crossloading from geometry to Item 1; furthermore, the size of the MI exceeded the total amount of misfit in the model, so this MI was disregarded. The addition of the covariances between Items 1 and 4, Items 7 and 11, Items 7 and 10, and Items 8 and 17 were also disregarded as theoretically weak because each pair of items measured different content knowledge and underlying misconception. Items 13 and 14, however, measured the same content knowledge, so their error covariance was considered theoretically plausible, and the parameter was added to the model. Table 58 displays the goodness of fit statistics for the resultant model (Model D4).

Table 58
Model D4 Goodness of Fit Indices, Covary Item 13 and 14 Errors

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	574.99	0.99	0.88	0.022	[0.000, 0.033]	0.074	0.65	0.731	1.36
1 (N = 577)	724.59	0.98	0.88	0.030	[0.021, 0.039]	0.074	0.65	0.724	1.32
2 (N = 553)	676.08	0.99	0.89	0.019	[0.000, 0.030]	0.066	0.66	0.734	1.14
3 (N = 558)	711.55	0.98	0.88	0.029	[0.019, 0.038]	0.072	0.65	0.724	1.35
4 (N = 566)	583.40	0.99	0.90	0.019	[0.000, 0.030]	0.070	0.66	0.726	1.13
5 (N = 575)	649.62	0.99	0.89	0.020	[0.0036, 0.031]	0.068	0.66	0.734	1.16
Wtd. Avg.	669.05	0.99	0.89	0.023	[0.014, 0.033]	0.070	0.66	0.728	1.22
SE_{Avg}	28.122	0.003	0.004	0.003		0.002	0.003	0.003	0.053

Note: *df* = 113

The difference of χ^2 test revealed a statistically significant decrease in model misfit in the imputed data sets ($\Delta\chi^2_{\text{Imputed Avg}}[1] = 8.194$, p = 0.004). The reduction in model misfit, however, was non-significant for the unimputed data ($\Delta\chi^2_{\text{Unimputed}}[1] = 0.79$, p = 0.374). Therefore, D4 was discarded; and as a result of no other theoretically reasonable MIs, Model D3 was retained as the best calibration of Model D for these data (Figure 29).

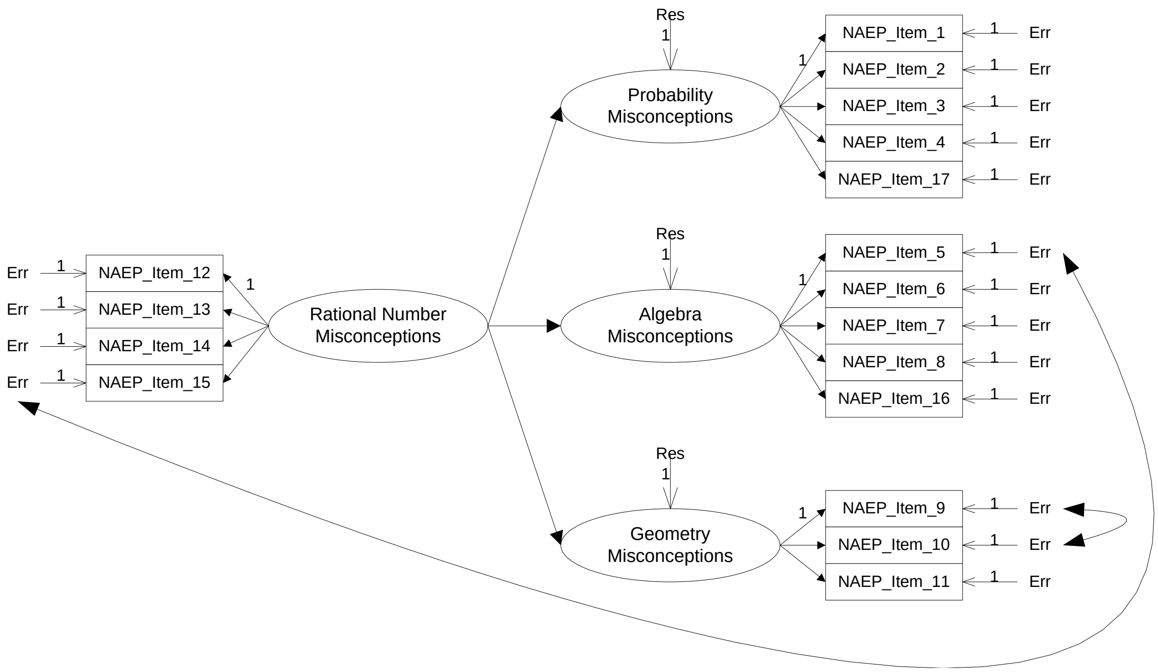


Figure 29. Final Structural Model D3.

Validation of Model D3. To examine the convergent validity of Model D3, the goodness of fit statistics were computed based on the validation sample (Table 59).

Table 59
Model D3 Goodness of Fit Indices from Validation Samples

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 449)	628.51	0.98	0.86	0.026	[0.011, 0.037]	0.081	0.64	0.729	1.51
1 (N = 556)	592.18	0.99	0.90	0.018	[0.000, 0.029]	0.068	0.67	0.737	1.12
2 (N = 580)	704.41	0.98	0.88	0.028	[0.017, 0.037]	0.074	0.66	0.735	1.25
3 (N = 575)	755.89	0.98	0.88	0.029	[0.019, 0.038]	0.077	0.66	0.735	1.28
4 (N = 567)	612.77	0.99	0.90	0.019	[0.000, 0.030]	0.067	0.67	0.737	1.09
5 (N = 558)	665.66	0.98	0.89	0.0267	[0.015, 0.035]	0.072	0.66	0.727	1.22
Wtd. Avg.	665.46	0.98	0.89	0.024	[0.015, 0.033]	0.072	0.66	0.734	1.19
SE_{Avg}	33.392	0.003	0.005	0.003		0.002	0.003	0.002	0.041

Note: *df* = 114

The goodness of fit statistics were then compared to those from the calibration sample using a *t*-test to compare the difference in the point estimates (Table 60).

Table 60
Model D3 Comparison of Calibration and Validation Sample Fit Indices

Imputation	χ^2	CFI	GFI	RMSEA	SRMR	PGFI	PCFI	ECVI
Imputed Data Set *t* value	0.42	0.68	-1.00	-0.03	-0.77	-1.00	-0.05	0.91

No statistic from the imputed data sets was significantly different for the calibration and validation samples (i.e., all t values less than 1.96), indicating that Model D3 fit the validation and calibration samples equally well. I, therefore, concluded that the model had good convergent validity across samples.

Analysis of Model E

Calibration. Model E reversed the relationship between probability and rational numbers from Model D, specifying misconceptions in probability as the independent variable with misconceptions in rational number, algebra, and geometry acting as dependent variables. The goodness of fit indices suggested that the Model E fit the calibration data very well (Table 61).

Table61
Model E Goodness of Fit Indices from Calibration Samples

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	606.76	0.99	0.87	0.024	[0.009, 0.035]	0.075	0.66	0.751	1.42
1 (N = 577)	751.43	0.98	0.87	0.031	[0.022, 0.040]	0.076	0.66	0.743	1.37
2 (N = 553)	680.20	0.99	0.89	0.018	[0.000, 0.029]	0.066	0.68	0.734	1.15
3 (N = 558)	778.03	0.98	0.87	0.032	[0.023, 0.041]	0.076	0.66	0.743	1.44
4 (N = 566)	646.52	0.99	0.88	0.025	[0.014, 0.034]	0.074	0.67	0.765	1.25
5 (N = 575)	716.60	0.99	0.88	0.025	[0.013, 0.034]	0.072	0.67	0.743	1.25
Wtd. Avg.	714.68	0.99	0.88	0.026	[0.017, 0.035]	0.073	0.67	0.746	1.29
SE_{Avg}	26.460	0.003	0.004	0.003		0.002	0.004	0.006	0.057

Note: *df* = 116

The maximum modification indices for Model E called for the addition of covariance parameters between the error terms for Items 5 and 15, Items 7 and 12, Items 6 and 12, and Items 13 and 14 (Table 62).

Table 62
Model E Maximum Modification Indices from Calibration Samples

Imputation	Maximum Modification Index	Associated Parameter to Add
0	33.17	TE(5, 4); Covariance between Item 5 and 15 errors
1	339.07	TE(7, 1); Covariance between Item 7 and 12 errors
2	42.79	TE(6, 1); Covariance between Item 6 and 12 errors
3	894.64	TE(6, 1); Covariance between Item 6 and 12 errors
4	68.22	TE(3, 2); Covariance between Item 13 and 14 errors
5	361.69	TE(5, 4); Covariance between Item 5 and 15 errors

The item pairs Items 6 and 12, Items 7 and 12, and Items 5 and 15 measured the same underlying misconception, so they were considered theoretically plausible. Item 13 and 14 measured the same content knowledge, so it was also considered theoretically plausible. Because the highest MI called for the addition of the error covariance between Items 6 and 12 and because that MI was called for by two data sets, I chose to disregard the fact that the highest MI was also larger than the total unaccounted variance in the model and added this parameter first (Table 63).

Table 63
Model E2 Goodness of Fit Indices from Calibration Samples, Covary Item 6 and 12 Errors

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	598.94	0.99	0.87	0.024	[0.0097, 0.035]	0.074	0.65	0.740	1.42
1 (N = 577)	729.62	0.98	0.87	0.031	[0.022, 0.040]	0.074	0.66	0.743	1.36
2 (N = 553)	633.47	0.99	0.90	0.017	[0.000, 0.028]	0.065	0.67	0.737	1.13
3 (N = 558)	761.00	0.98	0.87	0.032	[0.023, 0.041]	0.076	0.65	0.732	1.43
4 (N = 566)	632.07	0.99	0.89	0.024	[0.013, 0.034]	0.073	0.67	0.745	1.24
5 (N = 575)	704.59	0.99	0.88	0.024	[0.013, 0.034]	0.071	0.66	0.743	1.25
Wtd. Avg.	692.41	0.99	0.88	0.026		0.072	0.66	0.740	1.28
SE_{Avg}	28.888	0.003	0.007	0.003		0.002	0.004	0.003	0.058

Note: *df* = 115

The difference of χ^2 test revealed a statistically significant decrease in model misfit for both the unimputed data set (listwise deletion) and the imputed data sets ($\Delta\chi^2_{Unimputed}$ [1] = 7.82, p = 0.005; $\Delta\chi^2_{Imputed\ Avg}$[1] = 22.274, p < 0.0001). Therefore, Model E2 was retained, and the modification indices were examined (Table 64).

Table 64
Model E2 Maximum Modification Indices from Calibration Samples

Imputation	Maximum Modification Index	Associated Parameter to Add
0	36.16	TE(7, 6); Covariance between Item 6 and 7 errors
1	68.61	LY(5, 1); Crossloading, Rational Number to Item 5
2	42.79	TE(6, 1); Covariance between Item 6 and 12 errors
3	420.34	TE(7, 5); Covariance between Item 5 and 7 errors
4	67.67	TE(3, 2); Covariance between Item 13 and 14 errors
5	289.01	TE(5, 4); Covariance between Item 5 and 15 errors

The error covariance of Items 6 and 12 was rejected as theoretically implausible because the two items shared neither content area or underlying misconception. Items 13 and 14, Items 5 and 15, and Items 5 and 7 were considered theoretically plausible because each pair shared content area while measuring different underlying misconceptions. The error covariance between Items 6 and 7 offered the theoretically strongest MI; both items measured the same content area and underlying misconception. So, although the MI for this parameter was the smallest across the data sets, I chose to add it to the model next (Table 65).

Table 65
Model E3 Goodness of Fit Indices from Calibration Samples, Covary Item 6 and 7 Errors

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	592.29	0.99	0.87	0.024	[0.0084, 0.035]	0.074	0.65	0.740	1.41
1 (N = 577)	705.52	0.98	0.88	0.030	[0.021, 0.039]	0.074	0.65	0.724	1.34
2 (N = 553)	627.37	0.99	0.90	0.016	[0.000, 0.028]	0.065	0.67	0.737	1.12
3 (N = 558)	752.80	0.98	0.87	0.032	[0.023, 0.041]	0.075	0.65	0.732	1.42
4 (N = 566)	629.14	0.99	0.89	0.025	[0.013, 0.034]	0.073	0.66	0.734	1.23
5 (N = 575)	700.17	0.99	0.88	0.024	[0.013, 0.034]	0.071	0.66	0.743	1.24
Wtd. Avg.	683.20	0.99	0.88	0.025		0.072	0.66	0.734	1.27
SE_{Avg}	27.007	0.003	0.006	0.003		0.002	0.004	0.003	0.057

Note: *df* = 114

The difference of χ^2 test revealed a statistically significant decrease in model misfit for both the unimputed data set (listwise deletion) and the imputed data sets ($\Delta\chi^2_{Unimputed}[1] = 6.65$, $p = 0.010$; $\Delta\chi^2_{Imputed\ Avg}[1] = 9.21$, $p < 0.002$). Therefore, Model E3 was retained, and the modification indices were examined (Table 66).

Table 66
Model E3 Maximum Modification Indices from Calibration Samples

Imputation	Maximum Modification Index	Associated Parameter to Add
0	45.07	TE(12, 10); Covariance between Item 9 and 11 errors
1	81.17	LY(7, 1); Crossloading, Rational Number to Item 7
2	90.93	TE(12, 1); Covariance between Item 11 and 12 errors
3	1423.41	TE(5, 1); Covariance between Item 5 and 12 errors
4	131.15	TE(4, 1); Covariance between Item 12 and 15 errors
5	186.00	TE(5, 4); Covariance between Item 5 and 15 errors

The crossloading from rational number to Item 7 along with the error covariances of Items 5 and 12 and Items 11 and 12 were considered theoretically weak (i.e., no matching content knowledge or underlying misconception). The error covariances of Items 9 and 11 and Items 12 and 15 were considered theoretically plausible because each pair measured the same content knowledge or underlying misconception. The error covariance of Items 5 and 15 was considered the strongest plausible modification because each item measured rational number meaning misconceptions. Table 67 displays the goodness of fit indices for the model resulting from the addition of this parameter.

Table 67
Model E4 Goodness of Fit Indices from Calibration Samples, Covary Item 5 and 15 Errors

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	561.19	0.99	0.88	0.021	[0.000, 0.032]	0.073	0.65	0.731	1.34
1 (N = 577)	703.79	0.98	0.88	0.031	[0.021, 0.039]	0.074	0.65	0.724	1.33
2 (N = 553)	626.72	0.99	0.90	0.017	[0.000, 0.029]	0.064	0.66	0.726	1.12
3 (N = 558)	744.57	0.98	0.87	0.031	[0.021, 0.040]	0.075	0.64	0.721	1.39
4 (N = 566)	609.20	0.99	0.89	0.022	[0.0074, 0.032]	0.072	0.66	0.734	1.17
5 (N = 575)	661.25	0.99	0.89	0.021	[0.0055, 0.031]	0.069	0.66	0.734	1.17
Wtd. Avg.	669.20	0.99	0.89	0.024	[0.014, 0.035]	0.071	0.66	0.728	1.24
SE_{Avg}	27.777	0.003	0.006	0.003		0.002	0.004	0.003	0.058

Note: *df* = 113

The difference of χ^2 test revealed a statistically significant decrease in model misfit for both the unimputed data set (listwise deletion) and the imputed data sets ($\Delta\chi^2_{Unimputed}[1] = 31.1$, $p < 0.0001$; $\Delta\chi^2_{Imputed\ Avg}[1] = 14.003$, $p = 0.0002$). Therefore, Model E4 was retained, and the modification indices were examined (Table 68).

Table 68
Model E4 Maximum Modification Indices from Calibration Samples

Imputation	Maximum Modification Index	Associated Parameter to Add
0	32.06	TH(1, 11); Covariance between Item 1 and 10 errors
1	439.59	LY(7, 1); Crossloading, Rational Number to Item 7
2	83.13	TE(12, 1); Covariance between Item 11 and 12 errors
3	310.83	TH(2, 6); Covariance between Item 2 and 6 errors
4	76.70	TE(6, 5); Covariance between Item 5 and 6 errors
5	37.32	TH(5, 8); Covariance between Item 8 and 17 errors

None of the potential modifications to the model were theoretically plausible except for the error covariance between Items 5 and 6, which both measured algebra content knowledge. Therefore, the goodness of fit statistics for the resultant model were examined (Table 69).

Table 69

Model E5 Goodness of Fit Indices from Calibration Samples, Covary Item 5 and 6 Errors

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	545.31	0.99	0.88	0.020	[0.000, 0.032]	0.073	0.65	0.731	1.32
1 (N = 577)	695.07	0.98	0.88	0.031	[0.021, 0.039]	0.076	0.64	0.713	1.32
2 (N = 553)	602.29	0.99	0.90	0.016	[0.000, 0.028]	0.065	0.66	0.726	1.10
3 (N = 558)	736.66	0.98	0.87	0.031	[0.021, 0.040]	0.074	0.64	0.721	1.37
4 (N = 566)	587.68	0.99	0.89	0.021	[0.0041, 0.031]	0.071	0.65	0.723	1.15
5 (N = 575)	655.41	0.99	0.89	0.021	[0.0054, 0.031]	0.069	0.65	0.723	1.16
Wtd. Avg.	655.59	0.99	0.89	0.024	[0.013, 0.035]	0.071	0.65	0.721	1.22
SE_{Avg}	31.209	0.003	0.006	0.003		0.002	0.004	0.002	0.059

Note: *df* = 112

The difference of χ^2 test revealed a statistically significant decrease in model misfit for both the unimputed data set (listwise deletion) and the imputed data sets ($\Delta\chi^2_{Unimputed}[1] = 15.88$, $p < 0.0001$; $\Delta\chi^2_{Imputed\ Avg}[1] = 13.606$, $p = 0.0002$). Therefore, Model E5 was retained, and the modification indices were examined (Table 70).

Table 70

Model E5 Maximum Modification Indices from Calibration Samples

Imputation	Maximum Modification Index	Associated Parameter to Add
0	32.97	TH(1, 11); Covariance between Item 1 and 10 errors
1	56.44	TD(4, 1); Covariance between Item 1 and 4 errors
2	95.82	TE(11, 1); Covariance between Item 10 and 12 errors
3	83.26	TE(5, 1); Covariance between Item 5 and 12 errors
4	63.65	TE(3, 2); Covariance between Item 13 and 14 errors
5	37.17	TE(11, 10); Covariance between Item 9 and 10 errors

The error covariances between Items 1 and 10, Items 1 and 4, Items 10 and 12, and Items 5 and 12 were considered theoretically weak because the item pairs did not measure the same content knowledge or underlying misconception. The error covariance between Items 13 and 14 seemed theoretically plausible because both items measured rational number content knowledge, and both underlying misconceptions were related (i.e., absolute/relative comparison and rational number meaning misconceptions). The error covariance between Items 9 and 10 seemed the strongest theoretically because both items measured the same content area (geometry) and the same underlying misconception (spatial reasoning). The goodness of fit indices for the resultant model were therefore computed (Table 71).

Table 71
Model E6 Goodness of Fit Indices from Calibration Samples, Covary Item 9 and 10 Errors

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	542.33	0.99	0.88	0.020	[0.000, 0.032]	0.072	0.64	0.720	1.31
1 (N = 577)	663.77	0.98	0.88	0.028	[0.018, 0.037]	0.074	0.64	0.713	1.26
2 (N = 553)	597.31	0.99	0.90	0.016	[0.000, 0.028]	0.065	0.65	0.715	1.10
3 (N = 558)	713.58	0.98	0.88	0.030	[0.020, 0.039]	0.073	0.64	0.713	1.35
4 (N = 566)	579.89	0.99	0.90	0.020	[0.0024, 0.031]	0.070	0.65	0.715	1.13
5 (N = 575)	630.88	0.99	0.90	0.020	[0.000, 0.030]	0.068	0.65	0.715	1.13
Wtd. Avg.	637.14	0.99	0.89	0.023	[0.013, 0.033]	0.070	0.65	0.714	1.19
SE_{Avg}	26.757	0.003	0.005	0.003		0.002	0.003	0.001	0.053

Note: *df* = 111

The difference of χ^2 test revealed a statistically significant decrease in model misfit in the imputed data sets ($\Delta\chi^2_{\text{Imputed Avg}}[1] = 18.455$, $p < 0.0001$). The reduction in model misfit for the unimputed data, however, was only significant at the 91% confidence level ($\Delta\chi^2_{\text{Unimputed}}[1] = 2.98$, $p = 0.084$). I considered four criteria to determine that the model should be retained: (1) The unimputed data model was nearly significant and clearly no worse than the previous model, (2) The modification removed a large amount of misfit compared to the 95% CI critical value of four in the imputed model, (3) The expected cross validation values were smaller, so the addition of the parameter did not likely overfit to the sample, and (4) Three GFI values reached the recommended 0.90 cutoff despite non-normality. I therefore retained Model E6 and examined its modification indices (Table 72).

Table 72
Model E6 Maximum Modification Indices from Calibration Samples

Imputation	Maximum Modification Index	Associated Parameter to Add
0	45.00	TE(11, 1); Covariance between Item 10 and 12 errors
1	56.41	TD(4, 1); Covariance between Item 1 and 4 errors
2	284.54	TE(11, 1); Covariance between Item 10 and 12 errors
3	67.47	TD(4, 1); Covariance between Item 1 and 4 errors
4	62.24	TE(8, 6); Covariance between Item 6 and 8 errors
5	37.29	TH(5, 8); Covariance between Item 8 and 17 errors

The error covariances between Items 10 and 12, Items 1 and 4, and Items 8 and 17, were considered theoretically weak because the item pairs did not measure the same content knowledge or underlying misconception. The error covariance between Items 6 and 8 seemed theoretically plausible because both items measured algebra content knowledge, so the parameter was added to the model, and the goodness of fit indices for the resultant model were computed (Table 73).

Table 73
Model E7 Goodness of Fit Indices from Calibration Samples, Covary Item 6 and 8 Errors

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	527.60	0.99	0.88	0.020	[0.000, 0.032]	0.070	0.63	0.709	1.31
1 (N = 577)	648.02	0.98	0.89	0.028	[0.017, 0.037]	0.072	0.64	0.705	1.25
2 (N = 553)	595.06	0.99	0.90	0.016	[0.000, 0.028]	0.064	0.65	0.715	1.09
3 (N = 558)	697.44	0.98	0.88	0.031	[0.021, 0.040]	0.073	0.63	0.702	1.36
4 (N = 566)	540.98	0.99	0.90	0.019	[0.000, 0.030]	0.069	0.65	0.715	1.11
5 (N = 575)	618.79	0.99	0.90	0.020	[0.0027, 0.031]	0.067	0.64	0.704	1.14
Wtd. Avg.	620.06	0.99	0.89	0.023	[0.012, 0.033]	0.069	0.64	0.708	1.19
SE_{Avg}	29.208	0.003	0.004	0.003		0.002	0.004	0.003	0.057

Note: *df* = 110

The difference of χ^2 test revealed a statistically significant decrease in model misfit for both the unimputed data set (listwise deletion) and the imputed data sets ($\Delta\chi^2_{\text{Unimputed}}[1] = 14.73$, p = 0.0001; $\Delta\chi^2_{\text{Imputed Avg}}[1] = 17.077$, p < 0.0001). Therefore, Model E7 was retained, and the modification indices were examined (Table 74).

Table 74
Model E7 Maximum Modification Indices from Calibration Samples

Imputation	Maximum Modification Index	Associated Parameter to Add
0	58.10	TE(9, 6); Covariance between Item 6 and 16 errors
1	106.31	TE(6, 4); Covariance between Item 6 and 15 errors
2	167.49	TE(11, 1); Covariance between Item 10 and 12 errors
3	74.35	TE(5, 1); Covariance between Item 5 and 12 errors
4	64.56	TE(3, 2); Covariance between Item 13 and 14 errors
5	47.35	TH(5, 8); Covariance between Item 8 and 17 errors

The potential addition of the error covariances for Items 6 and 15, Items 10 and 12, Items 5 and 12, and Items 8 and 17 were discarded because each pair measured different content knowledge and underlying misconceptions. Items 13 and 14 and Items 6 and 16, on the other hand, measured the same content knowledge. I differentiated between the two error covariances by considering four characteristics: (1) The covariance of Items 13 and 14 had significantly reduced model misfit in previous models, (2) The underlying misconceptions for Items 13 and 14 were related (absolute/relative comparison and rational number meaning), (3) The MI for the error covariance of Items 13 and 14 was higher than for Items 6 and 16, and (4) The MI for the error covariance of Items 6 and 16 came from the unimputed data set. Since three of the four characteristics favored the addition of the error covariance for Items 13 and 14, this parameter was added to the model, and the goodness of fit indices were computed (Table 75).

Table 75
Model E8 Goodness of Fit Indices from Calibration Samples, Covary Item 13 and 14 Errors

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	526.90	0.99	0.88	0.020	[0.000, 0.032]	0.070	0.63	0.709	1.31
1 (N = 577)	645.67	0.98	0.89	0.028	[0.018, 0.037]	0.072	0.63	0.694	1.25
2 (N = 553)	594.35	0.99	0.90	0.017	[0.000, 0.029]	0.064	0.64	0.704	1.10
3 (N = 558)	696.54	0.98	0.88	0.031	[0.021, 0.040]	0.073	0.62	0.690	1.36
4 (N = 566)	506.12	1.00	0.90	0.015	[0.000, 0.027]	0.067	0.64	0.711	1.05
5 (N = 575)	618.69	0.99	0.90	0.021	[0.0048, 0.031]	0.067	0.64	0.704	1.15
Wtd. Avg.	612.27	0.99	0.89	0.022	[0.011, 0.034]	0.069	0.63	0.701	1.18
SE_{Avg}	35.220	0.004	0.004	0.003		0.002	0.004	0.004	0.062

Note: *df* = 109

The difference of χ^2 test revealed a statistically significant decrease in model misfit in the imputed data sets ($\Delta\chi^2_{\text{Imputed Avg}}[1] = 7.791$, p = 0.005). The reduction in model misfit, however, was non-significant for the unimputed data ($\Delta\chi^2_{\text{Unimputed}}[1] = 0.70$, p = 0.403). To reconcile this difference of significance, I considered that, although the imputed data showed a statistically significant change, only one of the data sets (Imputation 4) accounted for the change across all five data sets. Consequently, I discarded Model E8 and computed the goodness of fit indices for a re-adjusted Model E7 with its other theoretically reasonable MI, the error covariance of Items 6 and 16 (Table 76).

Table 76
Model E8ii Goodness of Fit Indices from Calibration Samples, Covary Item 6 and 16 Errors

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	523.83	0.99	0.88	0.020	[0.000, 0.032]	0.070	0.63	0.709	1.31
1 (N = 577)	647.05	0.98	0.89	0.028	[0.018, 0.037]	0.072	0.63	0.694	1.25
2 (N = 553)	590.05	0.99	0.90	0.016	[0.000, 0.028]	0.064	0.64	0.704	1.09
3 (N = 558)	693.63	0.98	0.88	0.031	[0.021, 0.040]	0.073	0.62	0.690	1.36
4 (N = 566)	537.89	0.99	0.90	0.019	[0.000, 0.030]	0.069	0.64	0.704	1.10
5 (N = 575)	609.35	0.99	0.90	0.020	[0.0028, 0.031]	0.066	0.64	0.704	1.14
Wtd. Avg.	615.59	0.99	0.89	0.023		0.069	0.63	0.699	1.19
SE_{Avg}	29.377	0.003	0.004	0.003		0.002	0.004	0.003	0.058

Note: *df* = 109

The difference of χ^2 test revealed a statistically significant decrease in model misfit in the imputed data sets ($\Delta\chi^2_{\text{Imputed Avg}}[1] = 4.466$, p = 0.035). The reduction in model misfit for the unimputed data, however, was only significant at the 93% confidence level ($\Delta\chi^2_{\text{Unimputed}}[1] = 3.77$, p = 0.035). Because these reductions were generally small across the data sets, and the additional parameter was theoretically marginal from the outset, I decided that Model E7 was the best calibration of Model E for these data (Figure 30).

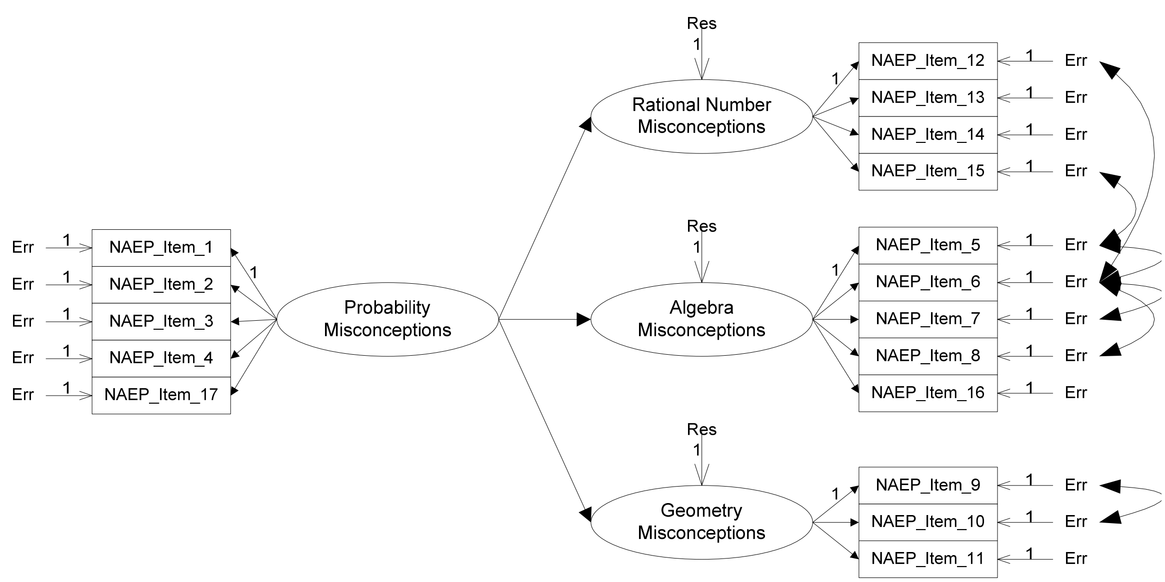


Figure 30. Final Structural Model E7.

Validation of Model E7. To examine the convergent validity of Model E7, the goodness of fit statistics were computed based on the validation sample (Table 77).

Table 77
Model E7 Goodness of Fit Indices from Validation Samples

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 449)	558.33	0.99	0.88	0.021	[0.000, 0.033]	0.078	0.63	0.709	1.40
1 (N = 556)	595.43	0.99	0.89	0.021	[0.0059, 0.032]	0.069	0.64	0.712	1.15
2 (N = 580)	670.25	0.98	0.89	0.028	[0.017, 0.037]	0.073	0.64	0.705	1.23
3 (N = 575)	727.03	0.98	0.89	0.029	[0.019, 0.038]	0.075	0.64	0.705	1.25
4 (N = 567)	635.27	0.99	0.90	0.023	[0.010, 0.033]	0.068	0.64	0.704	1.14
5 (N = 558)	652.34	0.98	0.89	0.027	[0.016, 0.036]	0.072	0.64	0.705	1.22
Wtd. Avg.	655.55	0.98	0.89	0.026	[0.020, 0.031]	0.071	0.64	0.706	1.20
SE_{Avg}	24.189	0.003	0.002	0.002		0.001	0.000	0.002	0.025

Note: *df* = 110

The goodness of fit statistics were then compared to those from the calibration sample using a *t*-test to compare the difference in the point estimates (Table 78).

Table 78
Model E7 Comparison of Calibration and Validation Sample Fit Indices

Imputation	χ^2	CFI	GFI	RMSEA	SRMR	PGFI	PCFI	ECVI
Imputed Data Set *t* value	-1.32	0.68	0.62	-1.05	-1.36	0.71	0.73	-0.17

No statistic from the imputed data sets was significantly different for the calibration and validation samples (i.e., all t values less than 1.96), indicating that Model D3 fit the validation and calibration samples equally well. I, therefore, concluded that the model had good convergent validity across samples.

Analysis of Model F

Calibration. Model F specified all four content area misconception factors as covarying independent variables (Figure 24F). The goodness of fit statistics indicated an excellent fit for the hypothesized model (Table 79).

Table 79
Model F Goodness of Fit Indices from Calibration Samples

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	602.38	0.99	0.87	0.025	[0.010, 0.036]	0.075	0.64	0.762	1.42
1 (N = 577)	734.17	0.97	0.87	0.033	[0.024, 0.041]	0.075	0.64	0.747	1.38
2 (N = 553)	667.59	0.99	0.90	0.018	[0.000, 0.029]	0.066	0.66	0.726	1.13
3 (N = 558)	722.26	0.98	0.87	0.030	[0.020, 0.039]	0.073	0.65	0.732	1.37
4 (N = 566)	632.29	0.98	0.88	0.025	[0.014, 0.035]	0.073	0.65	0.724	1.25
5 (N = 575)	694.49	0.99	0.89	0.024	[0.013, 0.034]	0.071	0.65	0.723	1.23
Wtd. Avg.	690.36	0.98	0.88	0.026	[0.017, 0.036]	0.072	0.65	0.730	1.27
SE_{Avg}	20.683	0.004	0.007	0.003		0.002	0.004	0.005	0.052

Note: *df* = 113

The maximum MIs from each data set were examined for potential parameters to reduce model misfit (Table 77). The largest MI pointed to a crossloading between rational number content and Item 7, an algebra item that measured additive/multiplicative structure misconceptions. Such a parameter seemed theoretically weak, and the magnitude of the MI (i.e., larger than the total χ^2 of each model) suggested that the source of the MI might be model instability rather than a substantive improvement in the model.

Table 80
Model F Maximum Modification Indices from Calibration Samples

Imputation	Maximum Modification Index	Associated Parameter to Add
0	142.40	TD(13, 1); Covariance between Item 5 and 12 errors
1	283.01	TD(11, 4); Covariance between Item 10 and 15 errors
2	51.97	TD(13, 12); Covariance between Item 5 and 11 errors
3	8339.60	LX(15, 1); Crossloading, Rational Number to Item 7
4	69.78	TD(3, 2); Covariance between Item 13 and 14 errors
5	145.14	TD(13, 12); Covariance between Item 5 and 11 errors

The error covariances for Items 5 and 12, Items 5 and 11, and Items 10 and 15 were considered theoretically weak additions to the model because each pair of items measured different content knowledge and different underlying misconceptions. The error covariance for Item 13 and 14 was considered theoretically plausible because the items measured the same content knowledge and related underlying misconceptions. So, I added the parameter to the model and computed the goodness of fit statistics (Table 81).

Table 81
Model F2 Goodness of Fit Indices from Calibration Samples

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 478)	601.99	0.99	0.87	0.025	[0.011, 0.036]	0.075	0.64	0.728	1.42
1 (N = 577)	731.23	0.97	0.87	0.033	[0.024, 0.041]	0.075	0.64	0.714	1.38
2 (N = 553)	666.89	0.99	0.90	0.018	[0.000, 0.030]	0.066	0.66	0.726	1.14
3 (N = 558)	721.38	0.98	0.87	0.031	[0.021, 0.040]	0.073	0.64	0.721	1.37
4 (N = 566)	598.51	0.98	0.89	0.023	[0.010, 0.033]	0.071	0.65	0.716	1.20
5 (N = 575)	694.45	0.99	0.89	0.025	[0.013, 0.034]	0.071	0.65	0.723	1.24
Wtd. Avg.	682.68	0.98	0.88	0.026	[0.016, 0.036]	0.071	0.65	0.720	1.27
SE_{Avg}	26.607	0.004	0.007	0.003		0.002	0.004	0.002	0.053

Note: *df* = 112

The difference of χ^2 test revealed a statistically significant decrease in model misfit in the imputed data sets ($\Delta\chi^2_{\text{Imputed Avg}}[1] = 8.000$, p = 0.005). The reduction in model misfit for the unimputed data, however, was not statistically significant ($\Delta\chi^2_{Unimputed}$ [1] = 0.39, p > 0.5). Most of the significant reduction in the imputed data sets occurred in the fourth data set while the rest of the data sets, including the unimputed data set, reflected no change in model misfit. Furthermore, the expected cross-validation statistic increased, which indicated that the new parameter may represent an overfitting of the model to a data set. Based on these considerations, I discarded Model F2 and returned to the original hypothesized model. Because none of the other MIs from the original model were theoretically plausible, I concluded that the original hypothesized model was the best calibration of Model F to these data (Figure 31).

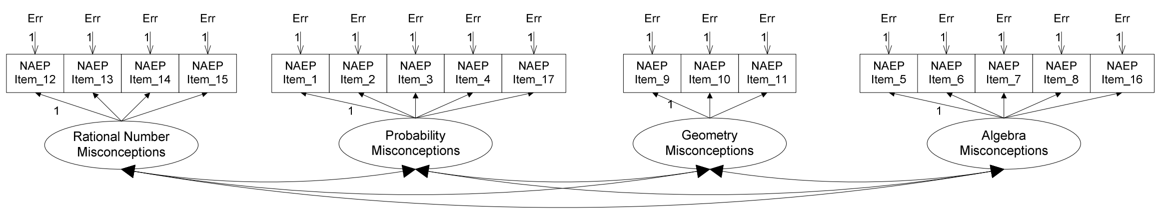


Figure 31. Final Structural Model F.

Validation of Model F. To examine the convergent validity of Model F, the goodness of fit statistics were computed based on the validation sample (Table 82).

Table 82
Model F Goodness of Fit Indices from Validation Samples

Imputation	χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
0 (N = 449)	633.59	0.98	0.86	0.029	[0.017, 0.040]	0.083	0.63	0.718	1.58
1 (N = 556)	626.94	0.99	0.89	0.023	[0.0090, 0.033]	0.071	0.66	0.734	1.18
2 (N = 580)	718.86	0.97	0.87	0.034	[0.025, 0.042]	0.078	0.64	0.714	1.39
3 (N = 575)	781.93	0.97	0.87	0.032	[0.023, 0.041]	0.078	0.65	0.725	1.36
4 (N = 567)	628.65	0.99	0.90	0.022	[0.0075, 0.032]	0.068	0.66	0.726	1.13
5 (N = 558)	685.22	0.98	0.88	0.028	[0.017, 0.037]	0.074	0.65	0.724	1.26
Wtd. Avg.	687.67	0.98	0.88	0.028	[0.019, 0.036]	0.074	0.65	0.725	1.26
SE_{Avg}	32.630	0.005	0.007	0.003		0.002	0.004	0.004	0.056

Note: *df* = 113

The goodness of fit statistics were then compared to those from the calibration sample using a *t*-test to compare the difference in the point estimates (Table 83).

Table 83
Model F Comparison of Calibration and Validation Sample Fit Indices

Imputation	χ^2	CFI	GFI	RMSEA	SRMR	PGFI	PCFI	ECVI
Imputed Data Set *t* value	0.10	0.44	0.00	-0.67	-1.00	-0.50	1.10	0.19

No statistic from the imputed data sets was significantly different for the calibration and validation samples (i.e., all t values less than 1.96), indicating that Model F fit the validation and calibration samples equally well. I, therefore, concluded that the model had good convergent validity across samples.

Summary of Structural Model Analysis

I originally hypothesized that either Models B or E would fit the data better than the others. The analysis of student response patterns indicated the possibility of a high degree of collinearity between the models: This collinearity resulted in every model fitting very well according to a wide range of fit indices and low parsimony as evidenced by parsimony indices higher than 0.5 (as recommended by Mulaik et al., 1989; Byrne, 2009). Table 84 summarizes the fit indices for the final calibration of each model.

Table 84
Summary of Fit Indices for Final Calibration of Each Model

Model	Statistic	χ^2	CFI	GFI	RMSEA	SRMR	PGFI	PCFI	ECVI
A2	Wtd. Avg.	701.59	0.98	0.88	0.025	0.072	0.664	0.741	1.26
	SE_{Avg}	28.967	0.004	0.005	0.003	0.002	0.004	0.003	0.058
B3	Wtd. Avg.	677.24	0.99	0.89	0.024	0.070	0.66	0.734	1.23
	SE_{Avg}	21.237	0.003	0.003	0.003	0.002	0.005	0.005	0.048
C3Aiii	Wtd. Avg.	682.82	0.99	0.89	0.024	0.071	0.66	0.732	1.23
	SE_{Avg}	22.736	0.003	0.003	0.002	0.002	0.004	0.004	0.045
C3Bii	Wtd. Avg.	703.00	0.98	0.88	0.025	0.072	0.66	0.741	1.25
	SE_{Avg}	27.783	0.004	0.005	0.003	0.002	0.004	0.003	0.052
D3	Wtd. Avg.	677.24	0.99	0.89	0.024	0.070	0.66	0.734	1.23
	SE_{Avg}	21.237	0.003	0.003	0.003	0.002	0.005	0.005	0.048
E7	Wtd. Avg.	620.06	0.99	0.89	0.023	0.069	0.64	0.708	1.19
	SE_{Avg}	29.208	0.003	0.004	0.003	0.002	0.004	0.003	0.057
F	Wtd. Avg.	687.67	0.98	0.88	0.028	0.074	0.65	0.725	1.26
	SE_{Avg}	32.630	0.005	0.007	0.003	0.002	0.004	0.004	0.056

These indices, while all excellent fit indices, are statistically indistinguishable across models. This result may be the result of a high degree of collinearity between content area misconceptions. The implications of this collinearity are discussed in Chapter 5.

Impact of Contextual Factors on Item Misconception Responses

Implementation of Intervention

Observations of classes and teacher interviews were used to assess fidelity of intervention implementation. Teachers from each school began at different times, usually upon completion of prior units. The pretest, ATMI, and MAI were administered by each teacher to their classes prior to the beginning of the treatment period. The treatment lasted between 5 and 10 class periods, depending on the teacher. The sample classes were observed in both the treatment and control conditions across all course types included in the study (Table 85).

Table 85
Observation Statistics for Fidelity of Implementation Checks

Class Grouping	Number of Classes	Number of Observations	Duration of Observations (Minutes)			
			Min	Median	Mean (SE)	Max
Total	53	42	20	20	27.14 (2.86)	90
Treatment	22	28	20	20	26.43 (3.68)	90
Control	28	14	20	20	28.57 (4.55)	60
Algebra 1	15	23	20	20	24.78 (3.44)	90
Geometry	17	8	20	20	30.00 (6.55)	60
Adv. Geometry	12	4	20	20	30.00 (10.00)	60
Algebra 2	4	4	20	20	37.50 (17.50)	90
Adv. Algebra 2	5	3	20	20	20.00 (0.00)	20

To maximize observation representativeness of treatment fidelity (i.e., concurrent criterion validity), days, times, and schedules of classroom visits were unannounced. These observations indicated that the intervention was not given to the control groups, nor were the treatment lessons interrupted with control group lessons.

Two Level Model

Two hierarchical analyses was conducted using HLM 6.08 (Raudenbush, Bryk, & Congdon, 2009) to examine the impact of item, student, and class characteristics on the emergence of errors due to mathematical misconceptions. The first analysis divided the model variance into two levels, student and class. The outcome variable for this model was the percent of misconception errors on the posttest. The second analysis divided th model variance into three levels, item, student, and class. The outcome variable for this model was a posttest misconception error indicator variable.

Descriptive statistics. Because of missing data in surveys, pretests, and posttests not accounted for by multiple imputation, samples sizes were different than those reported for other analyses. The observed sample sizes (Table 86) resulted in a statistical power of approximately 0.80 to detect a population effect size $\delta = 0.40$ and approximately 0.75 for a population effect size $\delta = 0.30$ for approximately 20 students per class. In this sample, class sizes averaged approximately 18 students.

Table 86

Descriptive Statistics for Two-Level HLM Model

Variable	N	Mean	SD
Student Level One			
PostPercent	567	0.35	0.18
PrePercent	567	0.38	0.17
Enjoyment	567	2.89	0.79
Motivation	567	2.87	0.93
Self Confidence	567	3.18	0.84
Value	567	3.49	0.76
Knowledge of Cognition	567	3.43	0.60
Regulation of Cognition	567	3.23	0.56
Class Level 2			
Class Mean Enjoyment	32	2.90	0.30
Class Mean Motivation	32	2.88	0.34
Class Mean Self Confidence	32	3.17	0.32
Class Mean Value	32	3.49	0.24
Class Mean Knowledge of Cognition	32	3.43	0.16
Class Mean Regulation of Cognition	32	3.23	0.16
Class Mean PrePercent	32	0.38	0.10

Unconditional null model. The unconditional ANOVA model (Equations 20 and 21) was examined first to determine the appropriateness of using a multilevel model to represent the data. The ANOVA HLM model, also referred to as the *null model*, was used to compute the intraclass correlation (ICC) and the overall mean for the dependent variable, percentage of misconceptions on the posttest (PostMis).

$$\text{Level 1} \qquad (20)$$
$$PostMis = \beta_0 + r_{ij}$$

$$\text{Level 2} \qquad (21)$$
$$\beta_0 = \gamma_{00} + u_{0j}$$

The variance for both levels (Table 87) was statistically significant at the 0.001 alpha level. The intraclass correlation was 0.229, meaning that 22.9% of the variance in the model is attributable to classroom effects.

Table 87
Unconditional Two Level Model Fixed and Random Effects

Fixed Effects	Coefficient	SE	T-Ratio	
Mean Posttest Misconceptions, γ_{00}	0.356	0.017	20.94	
Random Effects	*Variance Component*	*df*	χ^2	*p* Value
Between Classes, u_{0j}	0.0073	31	193.397	< 0.001
Within Classes, R	0.0245			

Using this null model as a baseline, the student model was developed to explain the impact of as many student characteristics as possible that may have been confounded by class effects (Ma, Ma, & Bradley, 2008).

Student Model. Using backward regression to develop the student model (Equations 22 and 23), all student level variables were entered into the null model.

Level 1 (22)

$$PostPerc_{ij} = \beta_0 + \beta_{1j}\left(Enjy_{ij} - \overline{Enjoy_{\bullet j}}\right) + \beta_{2j}\left(Mot_{ij} - \overline{Mot_{\bullet j}}\right)$$
$$+ \beta_{3j}\left(Slf\_Conf_{ij} - \overline{Slf\_Conf_{\bullet j}}\right) + \beta_{4j}\left(Value_{ij} - \overline{Value_{\bullet j}}\right) + \beta_{5j}\left(KCog_{ij} - \overline{KCog_{\bullet k}}\right)$$
$$+ \beta_{6j}\left(RCog_{ij} - \overline{RCog_{\bullet j}}\right) + \beta_{7j}\left(\text{NAEP\_Pre}_{ij} - \overline{\text{NAEP\_Pre}_{\bullet j}}\right) + r_{ij}$$

Level 2 (23)

$$\beta_0 = \gamma_{00} + u_{0j}$$
$$\beta_1 = \gamma_{10} + u_{1j}$$
$$\beta_2 = \gamma_{20} + u_{2j}$$
$$\beta_3 = \gamma_{30} + u_{3j}$$
$$\beta_4 = \gamma_{40} + u_{4j}$$
$$\beta_5 = \gamma_{50} + u_{5j}$$
$$\beta_6 = \gamma_{60} + u_{6j}$$
$$\beta_7 = \gamma_{70} + u_{7j}$$

Only pretest percentage of misconceptions and mathematics self confidence had a statistically significant effect on posttest percentage of misconceptions (Table 88).

Table 88
Student Characteristics Model Fixed and Random Effects

Fixed Effects	Coefficient	SE	T-Ratio	p Value
Mean Posttest Misconceptions, γ_{00}	0.356	0.017	21.34	< 0.001*
Enjoyment Slope, γ_{10}	0.019	0.014	1.35	0.188
Motivation Slope, γ_{20}	-0.006	0.010	-0.58	0.565
Self Confidence Slope, γ_{30}	-0.023	0.008	-2.71	0.011*
Value Slope, γ_{40}	-0.001	0.011	-0.110	0.914
Knowledge of Cognition Slope, γ_{50}	-0.022	0.017	-1.302	0.203
Regulation of Cognition Slope, γ_{60}	0.015	0.019	0.778	0.443
PrePercent Slope, γ_{70}	0.484	0.051	9.462	< 0.001*

Random Effects	*Variance Component*	*df*	χ^2	*p* Value
Mean Posttest Misconceptions, u_0	0.0077	31	277.98	< 0.001*
Enjoyment Slope, u_1	0.0009	31	25.47	> 0.500
Motivation Slope, u_2	0.0002	31	25.52	> 0.500
Self Confidence Slope, u_3	0.0002	31	21.19	> 0.500
Value Slope, u_4	0.0007	31	32.80	0.379
Knowledge of Cognition Slope, u_5	0.0008	31	31.72	0.430
Regulation of Cognition Slope, u_6	0.0022	31	35.94	0.248
PrePercent Slope, u_7	0.0256	31	37.41	0.198
Level 1, *R*	0.0171			

*Significant *p* values

The intercept of the student model represented a slightly different quantity than the null model: In the null model, γ_{00} represented the overall average percentage of misconceptions on the posttest; in the student model, γ_{00} represented the overall average percentage of misconceptions on the posttest after controlling for all level 1 predictors. So, γ_{00} represented the mean misconception percentage for a student who had an average score on pretest misconceptions; enjoyment, value, motivation, and self confidence; knowledge and regulation of cognition; and, unique student effects. The value of γ_{00} between the two models did not appear very different because only two of the fixed effects were statistically non-zero. Because of the non-significant fixed effects of most level 1 variables, only pretest misconception percentage and mathematics self confidence were retained for the contextual model.

The significance of the random effects in Table 88 provided two additional important pieces of information for the development of the contextual model. First, the only fixed effect with significant between-class variance to explain was the mean percentage of posttest misconceptions after controlling for all other level 1 variables. So, neither of the retained fixed effects were permitted to vary freely in Model 2. Second, the inclusion of the level 1 variables reduced the level 1 variance from 0.024 to 0.017, a 29% reduction. The remaining level 1 variance could not be explained by the other level 1 variables, so Model 2 (contextual model) left the level 1 variance untouched.

Contextual Model. The contextual model (Equations 24 and 25) began with the removal of all non-significant level 1 variables and non-significant level 2 random effects from the student model. Because no significant level 2 variance remained to be explained in the impact of the level 1 variables

Level 1 (24)

$$PostPerc_{ij} = \beta_0 + \beta_{1j}\left(Slf\_Conf_{ij} - \overline{Slf\_Conf_{\bullet j}}\right) + \beta_{2j}\left(\text{NAEP\_Pre}_{ij} - \overline{\text{NAEP\_Pre}_{\bullet j}}\right) + r_{ij}$$

Level 2 (25)

$$\begin{aligned}\beta_{0j} = \gamma_{00} &+ \gamma_{01}\left(Mean\_Enjoy_j - \overline{Mean\_Enjoy_{\bullet}}\right) + \gamma_{02}\left(Mean\_Mot_j - \overline{Mean\_Mot_{\bullet}}\right)\\ &+ \gamma_{03}\left(Mean\_SC_j - \overline{Mean\_SC_{\bullet}}\right) + \gamma_{04}\left(Mean\_Val_j - \overline{Mean\_Val_{\bullet}}\right)\\ &+ \gamma_{05}\left(Mean\_KCog_j - \overline{Mean\_KCog_{\bullet}}\right) + \gamma_{06}\left(Mean\_RCog_j - \overline{Mean\_RCog_{\bullet}}\right)\\ &+ \gamma_{07}\left(Mean\_Pre_j - \overline{Mean\_Pre_{\bullet}}\right) + \gamma_{08}(Treatment) + u_{0j}\end{aligned}$$

$$\beta_1 = \gamma_{10}$$
$$\beta_2 = \gamma_{20}$$

Table 89
Initial Contextual Model Fixed and Random Effects

Fixed Effects	Coefficient	SE	T-Ratio	p Value
For Intercept, β_0				
Mean Posttest Misconceptions, γ_{00}	0.360	0.007	52.66	< 0.001*
Mean Enjoyment, γ_{01}	-0.163	0.074	-2.19	0.038*
Mean Motivation, γ_{02}	-0.039	0.048	-0.82	0.423
Mean Self Confidence, γ_{03}	0.069	0.050	1.38	0.181
Mean Value, γ_{04}	0.108	0.051	2.13	0.044*
Mean Knowledge of Cognition, γ_{05}	-0.0004	0.088	-0.01	> 0.500
Mean Regulation of Cognition, γ_{06}	0.091	0.084	1.07	0.295
Mean PrePercent, γ_{07}	0.841	0.085	9.91	< 0.001*
Treatment, γ_{08}	-0.020	0.015	-1.307	0.202
For Self Confidence Slope, β_1				
Mean Self Confidence Slope, γ_{10}	-0.020	0.008	-2.57	0.011*
For PrePercent Slope, β_2				
Mean PrePercent Slope, γ_{20}	0.492	0.043	11.37	< 0.001*
Random Effect	*Variance Component*	*df*	χ^2	*p* Value
Mean Posttest Misconceptions, u_0	0.0003	24	41.30	< 0.001*
Level 1, *R*	0.0187			

*Significant *p* value

The initial contextual model (Table 89) revealed significant effects for the class mean mathematics enjoyment and value. The negative coefficient for enjoyment indicated that higher classroom levels of enjoyment of mathematics resulted in lower percentages of misconceptions on the posttest. Although the coefficient for value was positive, its magnitude was small enough that I hypothesized that it might be due to the non-significant variable inclusion in the model. This hypothesis was therefore tested in the final model. The unexplained between class variance was reduced 96% from 0.0077 to 0.0003, which resulted in a significant reduction in χ^2 ($\Delta\chi^2[7] = 236.68$, $p < 0.0001$). The signficant variables were retained for the final model.

Final Model. The final model (Equations 26 and 27) included only significant variables for both student level 1 and class level 2.

Student Level 1 (26)

$$PostPerc_{ij} = \beta_0 + \beta_{1j}\left(Slf\_Conf_{ij} - \overline{Slf\_Conf_{\bullet j}}\right) + \beta_{2j}\left(\text{NAEP\_Pre}_{ij} - \overline{\text{NAEP\_Pre}_{\bullet j}}\right) + r_{ij}$$

Class Level 2 (27)

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\left(Mean\_Enjoy_j - \overline{Mean\_Enjoy_{\bullet}}\right) + \gamma_{02}\left(Mean\_Val_j - \overline{Mean\_Val_{\bullet}}\right) + \gamma_{07}\left(Mean\_\text{Pre}_j - \overline{Mean\_\text{Pre}_{\bullet}}\right) + u_{0j}$$

$$\beta_1 = \gamma_{10}$$

$$\beta_2 = \gamma_{20}$$

Table 90
Final Contextual Model Fixed and Random Effects

Fixed Effects	Coefficient	SE	T-Ratio	p Value
For Intercept, β_0				
Class Mean Posttest Misconceptions, γ_{00}	0.359	0.007	48.639	< 0.001
Class Mean Enjoyment, γ_{01}	-0.129	0.041	-3.148	0.004
Class Mean Value, γ_{02}	0.132	0.053	2.492	0.019
Class Mean Pretest Misconceptions, γ_{03}	0.833	0.079	10.609	< 0.001
For Self Confidence Slope, β_1				
Grand Mean Self Confidence Slope, γ_{10}	-0.020	0.008	-2.57	0.011
For PrePercent Slope, β_2				
Grand Mean PrePercent Slope, γ_{20}	0.492	0.043	11.36	< 0.001
Random Effect	*Variance Component*	*df*	χ^2	*p* Value
Mean Posttest Misconceptions, u_0	0.007	29	221.60	< 0.001
Level 1, *R*	0.019			

The removal of non-significant variables from the contextual model (Table 90) added a significant amount of variance to the model, ($\Delta\chi^2[+5] = +180.30$, $p < 0.0001$). This result indicated that, although individual variables were non-significant, their cumulative effect may have been significant. One reason for this result may have been the reduced statistical power to detect smaller effect sizes. The reduction in χ^2 from the student model, however, was still statistically significant ($\Delta\chi^2[2] = 56.38$, $p < 0.0001$). The sign for class mean value remained positive, and the unexplained model variance increased significantly. I hypothesized that these effects may have been due to the removal of important cumulative effects of variables that were not significant by themselves.

To test this possibility, the two way interactions among the class mean value and class mean knowledge and regulation of cognition were computed by multiplying each the class mean value scores by the class mean knowledge and regulation of cognition scores (Pedhazur, 1997). To begin this investigation, the class mean value main effects were removed from the model to avoid multicollinearity. The interaction effects were then added to the class level 3 equation to produce a new model (Equations 28 and 29).

Student Level 1 (28)

$$PostPerc_{ij} = \beta_0 + \beta_{1j}\left(Slf\_Conf_{ij} - \overline{Slf\_Conf_{\bullet j}}\right) + \beta_{2j}\left(\text{NAEP\_Pre}_{ij} - \overline{\text{NAEP\_Pre}_{\bullet j}}\right) + r_{ij}$$

Class Level 2 (29)

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\left(Mean\_Enjoy_j - \overline{Mean\_Enjoy_{\bullet}}\right) + \gamma_{02}\left(Mean\_Pre_j - \overline{Mean\_Pre_{\bullet}}\right) + \gamma_{03}\left(Mn\_Val * KCog_j - \overline{Mn\_Val * KCog_{\bullet}}\right) + u_{0j}$$

$$\beta_1 = \gamma_{10}$$

$$\beta_2 = \gamma_{20}$$

Table 91

Post Hoc Model of Interaction Effects of Class Value of Mathematics and Knowledge of Cognition

Fixed Effects	Coefficient	SE	T-Ratio	p Value
For Intercept, β_0				
Class Mean Posttest Misconceptions, γ_{00}	0.359	0.007	50.075	< 0.001
Class Mean Enjoyment, γ_{01}	-0.141	0.040	-3.546	0.002
Class Mean Pretest Misconceptions, γ_{02}	0.848	0.077	10.952	< 0.001
Class Mean Value*Knowledge of Cognition Interaction, γ_{03}	0.031	0.011	2.920	0.007
For Self Confidence Slope, β_1				
Grand Mean Self Confidence Slope, γ_{10}	-0.020	0.008	-2.57	0.011
For PrePercent Slope, β_2				
Grand Mean PrePercent Slope, γ_{20}	0.492	0.043	11.36	< 0.001
Random Effect	*Variance Component*	*df*	χ^2	*p* Value
Mean Posttest Misconceptions, u_0	0.0005	28	46.39073	0.016
Level 1, *R*	0.019			

To interpret the interaction effect of mathematics value and knowledge of cognition on posttest misconceptions using coefficient values from Table 91 and means and standard deviations from Table 86, several predicted case values were examined, in which

γ_{00} = Mean posttest misconception percentage, controlling for all other variables in the model

γ_{01} = Impact of class ATMI enjoyment on posttest misconception percentage

γ_{02} = Impact of class pretest misconception percentage on student posttest misconception percentage

γ_{03} = Impact of interaction between class ATMI value and class MAI knowledge of cognition on student posttest misconception percentage

γ_{10} = Impact of self confidence on posttest misconception percentage

γ_{20} = Impact of pretest misconception percentage on posttest misconception percentage

1. A student who has an average mathematics self confidence (3.18) and a pretest misconception percentage that is average for the class, in a classroom with average knowledge of cognition (3.43), average classroom pretest misconception (0.38), and is 1 standard deviation (0.24 units) above the grand mean for value (3.49) is predicted by the post hoc model to have a posttest misconception score equal to the mean, 0.359.

Table 92
Predicted Value 1 for Two Level HLM Model

Coefficient	Value	Number of Units from Mean	Value Added
γ_{00}, Intercept	0.359	1	0.359
γ_{01}, Impact of Class ATMI Enjoyment	-0.141	0	0
γ_{02}, Impact of Class Pretest Misconception	0.848	0	0
γ_{03}, Impact of Class Value*Kcog	0.031	0.24*0 = 0	0
γ_{10}, Average Impact of Student Self Confidence	-0.02	0	0
γ_{20}, Average Impact of Student Pretest Misconception	0.492	0	0
PostPercent			0.359

Note: Value Added = (Coefficient Value) • (Number of Units from Mean)

Although the class value level in this example was 0.24 units above the grand mean, its interaction with knowledge of cognition negates its effect on the predicted posttest misconception error percentage. The second example shows an alternate effect of the interaction effect, when class knowledge of cognition is higher than the mean but value is equal to the mean.

2. A student who has an average mathematics self confidence (3.18) and a pretest misconception percentage that is average for the class, in a classroom with average mathematics value (3.49) and average classroom pretest misconception (0.38) and who is 1 standard deviation (0.16 units)

above the grand mean for knowledge of cognition (3.43) is predicted by the post hoc model to have a posttest misconception score of 0.359.

Table 93
Predicted Value 2 for Two Level HLM Model

Coefficient	Value	Number of units from average	Value Added
γ_{00}, Intercept	0.359	1	0.359
γ_{01}, Impact of Class ATMI Enjoyment	-0.141	0	0
γ_{02}, Impact of Class Pretest Misconception	0.848	0	0
γ_{03}, Impact of Class Value*Kcog	0.031	0*0.16 = 0	0
γ_{10}, Average Impact of Student Self Confidence	-0.02	0	0
γ_{20}, Average Impact of Student Pretest Misconception	0.492	0	0
PostPercent			0.359

Note: Value Added = (Coefficient Value) • (Number of Units from Mean)

Just as in the first predicted value (Table 92), the interaction of class value and knowledge of cognition eliminates the effect of knowledge of cognition on the posttest misconception error percentage. The third predicted value shows the effect of class value and knowledge of cognition when neither variable is equal to its grand mean.

3. A student who has an average mathematics self confidence (3.18) and a pretest misconception percentage that is average for the class, in a classroom with average classroom pretest misconceptions (0.38) and is 1 standard deviation (0.16 units) above the grand mean for knowledge of cognition (3.43) and 1 standard deviation (0.24 units) above the mean of value is predicted by the post hoc model to have a posttest misconception score of 0.360.

Table 94
Predicted Value 3 for Two Level HLM Model

Coefficient	Value	Number of units from average	Value Added
γ_{00}, Intercept	0.359	1	0.359
γ_{01}, Impact of Class ATMI Enjoyment	-0.141	0	0
γ_{02}, Impact of Class Pretest Misconception	0.848	0	0
γ_{03}, Impact of Class Value*Kcog	0.031	0.24*0.16 = 0.0384	0.0012
γ_{10}, Average Impact of Student Self Confidence	-0.02	0	0
γ_{20}, Average Impact of Student Pretest Misconception	0.492	0	0
PostPercent			0.3602

Note: Value Added = (Coefficient Value) • (Number of Units from Mean)

Although both class value and knowledge of cognition are above their grand means, their combined effect only increased the predicted percentage of misconception errors by 1%. The fourth predicted value shows the effect of student mathematics self confidence in a class with low mathematics value but high knowledge of cognition.

4. A student who has a mathematics self confidence 1 standard deviation (0.84 units) above the mean (3.18) and a pretest misconception percentage that is average for the class, in a classroom with average pretest misconceptions (0.38) and 1 standard deviation (0.24) below the mean of value (3.49) and 1 standard deviation (0.16) above the mean of knowledge of cognition is predicted by the post hoc model to have a posttest misconception score equal to 0.341.

Table 95
Predicted Value 4 for Two Level HLM Model

Coefficient	Value	Number of units from average	Value Added
γ_{00}, Intercept	0.359	1	0.359
γ_{01}, Impact of Class ATMI Enjoyment	-0.141	0	0
γ_{02}, Impact of Class Pretest Misconception	0.848	0	0
γ_{03}, Impact of Class Value*Kcog	0.031	-0.24*0.16 = -0.0384	-0.0012
γ_{10}, Average Impact of Student Self Confidence	-0.020	0.84	-0.0168
γ_{20}, Average Impact of Student Pretest Misconception	0.492	0	0
PostPercent			0.341

Note: Value Added = (Coefficient Value) • (Number of Units from Mean)

The effect of student self confidence was greater than the effect of the value-knowledge of cognition interaction even though the coefficient had a smaller magnitude because of the relative sizes of the standard deviation; the student self confidence standard deviation was almost four times larger than value and five times larger than for knowledge of cognition. The final predicted value example for this model shows the effect of reversing the relative class position for value and knowledge of cognition with respect to their grand means.

5. A student who has a mathematics self confidence 1 standard deviation (0.84 units) above the mean (3.18) and a pretest misconception percentage that is average for the class, in a classroom with average pretest misconceptions (0.38) and 1 standard deviation (0.24 units) above the mean of value (3.49) and 1 standard deviation (0.16) below the mean of knowledge of cognition is predicted by the post hoc model to have a posttest misconception score equal to 0.341.

Table 96
Predicted Value 5 for Two Level HLM Model

Coefficient	Value	Number of units from average	Value Added
γ_{00}, Intercept	0.359	1	0.359
γ_{01}, Impact of Class ATMI Enjoyment	-0.141	0	0
γ_{02}, Impact of Class Pretest Misconception	0.848	0	0
γ_{03}, Impact of Class Value*Kcog	0.031	0.24*-0.16 = -0.0384	-0.0012
γ_{10}, Average Impact of Student Self Confidence	-0.020	0.84	-0.0168
γ_{20}, Average Impact of Student Pretest Misconception	0.492	0	0
PostPercent			0.341

Note: Value Added = (Coefficient Value) • (Number of Units from Mean)

In the fourth example, class value was one standard deviation below while knowledge of cognition was one standard deviation above their means. In this example, their position from their grand means is reversed. This change resulted in no change to the percentage of misconception errors predicted by the model.

Three Level Bernoulli Model

The student level of the HLM model was then divided into two levels, item characteristics and student characteristics. By doing so, the outcome variable become a dichotomous variable representing a misconception error for each item for each student in each class. The new model was then examined using a generalized HLM model (HGLM) to measure the probability of misconception errors. The initial null model was examined to determine the amount of variance at each level: item level 1, student level 2, and class level 3.The contextual model was then used to evaluate the impact of each variable on the outcome.

Descriptive Statistics. The observed sample sizes (Table 97) resulted in a statistical power of approximately 0.80 to detect a population effect size $\delta = 0.40$ and approximately 0.75 for a population effect size δ = 0.30 for approximately 20 students per class. In this sample, class sizes averaged approximately 18 students.

Table 97

Descriptive Statistics for Three-Level HLM

Variable	N	Mean	SD	Min	Max
Item Level One					
Misconception	9673	0.35	0.48	0.00	1.00
Discrimination	9673	0.80	0.20	0.44	1.22
Difficulty	9673	0.01	0.51	-1.21	0.96
Moderate	9673	0.35	0.48	0.00	1.00
Student Level 2					
Enjoyment	515	2.91	0.79	1.00	5.00
Motivation	515	2.90	0.93	1.00	5.00
Self Confidence	515	3.19	0.83	1.00	5.00
Value	515	3.51	0.75	1.00	5.00
Knowledge of Cognition	515	3.43	0.58	1.18	4.94
Regulation of Cognition	515	3.23	0.56	1.11	4.74
NAEP_Pretest Percent Misconception	515	0.38	0.17	0.00	0.76
Class Level 3					
Mean Enjoyment	32	2.90	0.30	2.35	3.52
Mean Motivation	32	3.43	0.16	3.01	3.65
Mean Self Confidence	32	3.23	0.16	2.92	3.50
Mean Value	32	0.38	0.10	0.14	0.53
Mean Knowledge of Cognition	32	2.90	0.30	2.35	3.52
Mean Regulation of Cognition	32	2.88	0.34	2.15	3.63
Mean Pretest Percent Misconception	32	3.17	0.32	2.41	3.96
Treatment	32	0.50	0.51	0.00	1.00

Unconditional Null Model. The unconditional model (Equations 30, 31, and 32) revealed a significant amount of variance at both the student Level 2 and class Level 3 (Table 98). Additionally, the level 1 variance was also statistically significant (SE = 0.014, t = 67.71)

Item Level 1 (30)

$$Prob(Misconception_{ijk} = 1 \mid \pi_{jk}) = \varphi_{ijk}$$

$$Log\left[\frac{\varphi_{ijk}}{1-\varphi_{ijk}}\right] = \eta_{ijk}$$

$$\eta_{ijk} = \pi_{0jk} + e_{ijk}$$

Student Level 2 (31)

$$\pi_{0jk} = \beta_{0k} + r_{0jk}$$

Class Level 3 (32)

$$\beta_0 = \gamma_{00} + u_{0k}$$

Table 98

Unconditional Three Level Model Fixed and Random Coefficients

Fixed Effects	Logit Link: Unit-Specific Model		Logit Link: Population Average Model	
Mean Item Misconception, γ_{000}	-0.643**		-0.590**	
Random Effects	*Variance Component*	*df*	χ^2	*p* Value
Between Classes, u_{00}	0.184**	31	188.001	< 0.001
Between Students, R_0	0.299**	483	1117.002	< 0.001
Between Items, E	0.945**			

*Indicates |coeff/se| > 2.00; ** Indicates |coeff/se| > 3.00;

The outcome variable for Level 1 is in logit units, or the natural logarithm of the odds ratio, as shown in Equation 30. The coefficients, therefore, are also computed in logit units. Using the logit unit, the relationship between the outcome variable and independent variable coefficients have a linear relationship. Once the coefficient logit is converted to a probability, its relationship to the outcome variable and other logit coefficients is no longer linear. Therefore, to compute a predicted probability of misconception error for an item, the predicted logit value must be computed first. Conversion to a probability is the final step in predicting outcomes in the Bernoulli HLM model. The process of converting from a logit to a probability requires two steps. First, the logit is converted to an odds ratio using Equation 33.

$$\text{Odds} = e^{\text{Logit}} \tag{33}$$

Second, the inverse of the odds ratio is used to compute the probability using Equation 34.

$$\text{Probability} = \left(\frac{1}{1 + e^{-Logit}} \right) \tag{34}$$

The γ_{000} logit (unit specific model) of -0.643 (Table 93) corresponds to an odds ratio of 0.526, or a probability of 0.35 for the appearance of a misconception on an item, which corresponds to the mean for Item Misconception (see Table 97). The logit of the population average model (-0.590) indicates that the expected appearance of misconceptions in the population is slightly different from the observed sample misconception probability, an odds ratio of 0.554 and a probability of 0.357. This difference represents the expected effect of τ_{00}, in this case pulling the mean value of the unit specific model upward toward a probability of 0.50 (Raudenbush & Bryk, 2002).

The total variance in the model equals the sum of the variance from all three levels (Table 98), 0.184 + 0.299 + 0.945 = 1.428. The proportion of variance at the item Level 1 is 0.184/1.428 = 0.129 = 12.9%.. The proportion of variance at the student Level 2 is 0.299/1.428 = 0.209 = 20.9%. The proportion of variance at the class Level 3 is 0.945/1.428 = 0.662 = 66.2%. In summary, the variance at each level was statistically significant, and the class level 3 accounted for the majority of the variance in the probability of misconception errors. To begin accounting for variance, the item level 1 model was calibrated.

Item level 1 model. The discrimination and difficulty IRT coefficients for each NAEP item (see Table 10)were used as explanatory variables in the level 1 model. Additionally, the reported level of complexity assigned by NAEP reviewers (see Table 8) was added as a dichotomous predictor of misconception errors on a particular item (Low Complexity = 0; Moderate Complextiy = 1). Two models were examined before arriving at the final item model (Equations 35, 36, and 37; Table 99).

Item Level 1 (35)

$$Prob\left(Misconception_{ijk} = 1 \mid \pi_{jk}\right) = \varphi_{ijk}$$

$$Log\left[\frac{\varphi_{ijk}}{1-\varphi_{ijk}}\right] = \eta_{ijk}$$

$$\eta_{ijk} = \pi_{0jk} + \pi_{1jk}\left(Discrimination\right) + \pi_{2jk}\left(Difficulty\right) + \pi_{3jk}\left(Complexity\right) + e_{ijk}$$

Student Level 2 (36)

$$\pi_{0jk} = \beta_{00k} + r_{0jk}$$
$$\pi_{1jk} = \beta_{10k} + r_{1jk}$$
$$\pi_{2jk} = \beta_{20k} + r_{2jk}$$
$$\pi_{3jk} = \beta_{30k}$$

Class Level 3 (37)

$$\beta_{00k} = \gamma_{000} + u_{00k}$$
$$\beta_{10k} = \gamma_{100}$$
$$\beta_{20k} = \gamma_{200} + u_{20k}$$
$$\beta_{30k} = \gamma_{300}$$

The variance components for class level discrimination (U10), class level complexity (U30), and student level complexity (R3) were statistically non-significant, so they were fixed in the final item model. The addition of the discrimination, difficulty, and complexity variables reduced the item level variance from 0.945 to 0.881, a 6.8% reduction.

Table 99
Item Model of Mathematics Misconception Errors

	Unit Specific Model		Population Average Model	
Fixed Effects	Unconditional Model[b]	Final Model	Unconditional Model[b]	Final Model
Intercept, γ_{000}	-0.643**	0.233*	-0.590	0.130
Discrimination, γ_{100}	—	-1.206***	—	-0.957***
Difficulty, γ_{200}	—	0.682***	—	0.565***
Complexity, γ_{300}	—	0.080	—	0.094*

Variance Components	*Unconditional Model[b]*	*Final Model*
Intercept U_{00}, $\tau(\beta)_{11}$	0.184***	0.195***
Class Discrimination U_{10}, $\tau(\beta)_{22}$	—	—
Class Difficulty, U_{20}, $\tau(\beta)_{33}$	—	0.056*
Class Complexity, U_{30}, $\tau(\beta)_{44}$	—	—
Std Intercept, R_0, $\tau(\pi)_{11}$	0.299***	1.044***
Std Discrimination, R_1, $\tau(\pi)_{22}$	—	2.480***
Std Difficulty, R_2, $\tau(\pi)_{33}$	—	0.385***
Std Complexity, R_3, $\tau(\pi)_{44}$	—	—
Item, E, σ^2	0.945[a]	0.881[a]

*p < 0.05, **p < 0.01, ***p < 0.001, [a]|coeff /SE| > 3.00
[b]$df_{Student} = 483$, $df_{Class} = 31$

Using equations 33 and 34, the logits for the final item model were converted into predicted probabilities of misconception errors for different item characteristics. The intercept logit value predicts that the probability of a misconception error for a non-discriminating item (i.e., item characteristic curve = horizontal line) of average difficulty (i.e., Difficulty = 0) and low complexity is 0.558 in the sample and 0.532 in the population (Table 100). If the difficulty of a non-discriminating item increases difficulty by one standard deviation (0.62), then the probability of a misconception error increases to 0.658 for the sample and 0.618 for the population (Table 100).

If an item has average discrimination (0.825, mean of discrimination values from Table 10), then the predicted probability of a misconception error reduces to 0.318 in the sample and 0.341 in the population. If the discrimination of an item has a value one standard deviation above the average discrimination (0.8 + 0.2 = 1), then the predicted probability of a misconception error reduces to 0.230 in the sample and 0.277 in the population (Table 100). If a non-discriminating item with an average difficulty level increases from low to moderate complexity, the probability of a misconception error inceases to 0.578 for the sample and 0.556 for the population (Table 100).

Table 100
Selected Predicted Values for Final Item Model

	Unit Specific Model			Population Average Model		
Fixed Effects	Logit	Odds	Probability	Logit	Odds	Probability
INTERCEPT, γ_{000}	0.234	1.263	0.558	0.130	1.139	0.532
Discrimination, γ_{100}	-1.206	0.299	0.230	-0.957	0.384	0.277
Difficulty, γ_{200}	0.682	1.978	0.664	0.565	1.759	0.638
Complexity, γ_{300}	0.080	1.083	0.520	0.094	1.099	0.523
Combined Effects						
Int + Mean Discrimination	-0.731	0.481	0.325	-0.636	0.530	0.346
Int + Complexity	0.314	1.368	0.578	0.224	1.251	0.556
Int + 2SD Above Mean Discrimination	-1.696	0.183	0.155	-1.018	0.361	0.265
Int + 1SD Above Mean Difficulty	0.588	1.801	0.643	0.424	1.528	0.604
Int + 1SD Below Mean Difficulty	-0.107	0.898	0.473	-0.153	0.859	0.462

The probabilities and odds ratios for each logit value in Table 100 were computed using Equations 33 and 34. Combined effects were computed through a process of three steps. First, standard deviations of discrimination and difficulty were taken from Table 97. Second, the relevant number of standard deviations values were multiplied by the logit coefficient and added to the intercept logit. Third, the resulting logit sum was converted to an odds ratio and probability using Equations 33 and 34.

These predicted probability values reflect a statistically significant impact of item characteristics on the probability of a misconception error. The remaining variance of the item level was still statistically significant after the addition of all available variables, indicating that a future examination of other item characteristics may be beneficial to understanding item characteristic influences on misconception errors. The final item model was used as the starting point for calibration of the student model.

Student level 2 model. Student characteristics were added to level 2 (student level) of the final item model to examine the impact of student characteristics on the probability of a misconception error and on the impact of item characteristics on the probability of a misconception error. Only statistically significant effects were retained in the final model (Table 101) with the exception of the self confidence impact on the difficulty slope. Self confidence was retained because removing it from the model resulted in the loss of a significant coefficient for motivation, which was statistically significant in all intermediate models.

Table 101
Final Student Model Fixed and Random Coefficients

Fixed Effects	Logit Link: Unit-Specific Model	Logit Link: Population Average Model
Intercept, γ_{000}	0.237	0.165
Discrimination Slope, γ_{100}	-1.221***	-1.027***
Pretest Slope, γ_{110}	3.389***	2.961***
Difficulty Slope, γ_{200}	0.689***	0.583***
Motivation Slope, γ_{210}	-0.199*	-0.182*
Self Confidence Slope, γ_{220}	0.163	0.151
Pretest Slope, γ_{230}	-0.799*	-0.525
Complexity Slope, γ_{300}	0.080	0.095*

Random Effects	*Variance Component*	*df*	χ^2	*p* Value
Class Intercept, u_{00}	0.214***	31	221.117	< 0.001
Class Difficulty Slope	0.058***	31	53.032	0.008
Std Intercept, R_0	1.075***	483	622.784	< 0.001
Std Discrimination Slope, R_1	2.265***	513	686.892	< 0.001
Std Difficulty Slope, R_2	0.348***	480	649.958	< 0.001
Item Intercept, E	0.887***			

*p < 0.05, **p < 0.01, ***p < 0.001, [a]|coeff /SE| > 3.00

As with the Item Model, predicted values are presented to clarify the meaning of coefficients computed as logits (Table 102). The process for computing these predicted probabilities was the same as for the Item Model (i.e., use Equations 33 and 34 to convert logits to probability). The fixed effect intercept, γ_{000}, represents a predicted probability of a misconception error on an item with no discrimination, average difficulty, and low complexity of 0.559 for the sample and 0.541 in the population. The fixed effect for discrimination, γ_{100}, means that the impact of a change of one logit unit in discrimination (for an item with average difficulty and low complexity) corresponds to a probability of misconception error of 0.228 for the sample and 0.264.

Table 102
Selected Predicted Values for Final Student Model

	Unit Specific Model			Population Average Model		
Fixed Effects	Logit	Odds	Probability	Logit	Odds	Probability
Intercept, γ_{000}	0.237	1.267	0.559	0.165	1.179	0.541
Discrimination Slope, γ_{100}	-1.221	0.295	0.228	-1.027	0.358	0.264
Pretest Slope, γ_{110}	3.389	29.636	0.967	2.961	19.317	0.951
Difficulty Slope, γ_{200}	0.689	1.992	0.666	0.583	1.791	0.642
Motivation Slope, γ_{210}	-0.199	0.820	0.450	-0.182	0.834	0.455
Self Confidence Slope, γ_{220}	0.163	1.177	0.541	0.151	1.163	0.538
Pretest Slope, γ_{230}	-0.799	0.450	0.310	-0.525	0.592	0.372
Complexity Slope, γ_{300}	0.08	1.083	0.520	0.095	1.100	0.524
Combined Effects	$\hat{\eta}$	Odds	Probability	$\hat{\eta}$	Odds	Probability
Int + Mean Discrimination + Mean Difficulty + Low Complexity + Mean Pretest	-0.894	0.409	0.290	-0.862	0.422	0.297
Int + Mean Discrimination + Mean Difficulty + Low Complexity + 1SD Above Mean Pretest	-0.279	0.757	0.431	-0.254	0.776	0.437
Int + Mean Discrimination + 1SD Below Mean Difficulty + 1SD Above Mean Motivation + 1SD Above Mean Pretest	-0.174	0.840	0.457	-0.152	0.859	0.462
Int + Mean Discrimination + 1SD Above Mean Difficulty + 1SD Below Mean Motivation + 1SD Above Mean Pretest	0.677	1.968	0.663	0.613	1.846	0.649

The combined effects in Table 102 were computed by adding the relevant student logits to the item logits, then combining the item logits to produce the predicted logit for misconception errors ($\hat{\eta}$). These combined logits were then converted to probabilities using Equations 33 and 34. If a student with more pretest misconceptions (1 SD = 0.17) than the mean (0.38) completed an item with average discrimination (0.8), average difficulty (0), and low complexity, the probability of a misconception error increases to

0.431 in the sample and 0.437 in the population. If a student with more misconceptions (1 SD = 0.17) and more motivation (1 SD = 0.93) than the mean (0.38, 2.9 respectively) completed an easy item (1 SD below difficulty mean = -0.5), the probability of a misconception error was 0.457 in the sample and 0.462 in the population. If a student with more misconceptions (1SD = 0.17) and less motivation (1 SD = 0.93) than the mean (0.38, 2.9 respectively) completed a difficult item (1 SD above difficulty mean = 0.502) with mean discrimination (0.8), the probability of a misconception error increases to 0.663 in the sample and 0.649 in the population.

Two student level slopes demonstrated statistically significant variance at Level 3 (class), the intercept U000 and motivation slope U200. The final student model was used as the initial model for calibrating class level variables.

Class level 3 model. Class characteristics were added as predictors to the two level 3 equations with statistically significant, the intercept (mean probability of misconception error) and the difficulty slope (impact of item difficulty on the probability of misconception error).

The addition of these class parameters produced the final model (Table 103), which reduced the class variance from 0.214 to 0.002, a 98.7% reduction. This reduction reflected a statistically significant reduction in level 3 model misfit ($\Delta\chi^2 = 188.482$, $\Delta df = 8$, $p < 0.001$).

Table 103
Final Class Model Fixed and Random Coefficients

Fixed Effects	Logit Link: Unit-Specific Model	Logit Link: Population Average Model
Intercept, γ_{000}	0.268*	0.200[b]
Class Enjoyment, γ_{001}	-0.885*	-0.867*
Class Motivation, γ_{002}	-0.109	-0.079
Class Self Confidence, γ_{003}	0.234	0.246
Class Value, γ_{004}	0.649*	0.597*
Class Knowledge of Cognition, γ_{005}	0.118	0.164
Class Regulation of Cognition, γ_{006}	0.349	0.307
Class Pretest, γ_{007}	4.676***	4.423***
Treatment, γ_{008}	-0.051	-0.055
Discrimination Slope, γ_{100}	-1.215***	-1.067***
Pretest Slope, γ_{110}	3.379***	3.100***
Difficulty Slope, γ_{200}	0.814***	0.730***
Class Enjoyment, γ_{201}	-1.384*	-1.319*
Class Motivation, γ_{202}	0.338	0.290
Class Self Confidence, γ_{203}	0.961*	0.894*
Class Value, γ_{204}	0.118	0.158
Class Knowledge of Cognition, γ_{205}	0.138	0.090
Class Regulation of Cognition, γ_{206}	0.350	0.363
Class Pretest, γ_{207}	-0.536	-0.257
Treatment, γ_{208}	-0.231[b]	-0.208[b]
Motivation Slope, γ_{210}	-0.199*	-0.187*
Self Confidence Slope, γ_{220}	0.162	0.154
Pretest Slope, γ_{230}	-0.792*	-0.583
Complexity Slope, γ_{300}	0.080	0.097*

Random Effects	*Variance Component*	*df*	χ^2	*p* Value
Class Intercept, u_{00}	0.002[b]	23	32.635	0.088
Class Difficulty Slope, u_{20}	0.008*	23	35.125	0.050
Std Intercept, R_0	1.044***	483	621.070	< 0.001
Std Discrimination Slope, R_1	1.494***	513	683.774	< 0.001
Std Difficulty Slope, R_2	0.597***	480	650.180	< 0.001
Item Intercept, E	0.943[a]			

$*p < 0.05$, $**p < 0.01$, $***p < 0.001$, [a]$|coeff/SE| > 3.00$, [b]$p \leq 0.10$

The treatment condition was not a statistically significant predictor of the intercept (mean probability of misconception error), but it was a statistically significant predictor of the difficulty slope at the 90% confidence level for both the sample and population models. The coefficients from Table 103 were used to compute the predicted probability for a misconception error under various item, student, and class conditions (Table 104).

Table 104

Selected Predicted Values for Final Class Model

	Unit Specific Model			Population Average Model		
Fixed Effects	Logit	Odds	Probability	Logit	Odds	Probability
Intercept, γ_{000}	0.268	1.307	0.567	0.200	1.221	0.550
Class Enjoyment, γ_{001}	-0.885	0.413	0.292	-0.867	0.420	0.296
Class Motivation, γ_{002}	-0.109	0.897	0.473	-0.079	0.924	0.480
Class Self Confidence, γ_{003}	0.234	1.264	0.558	0.246	1.279	0.561
Class Value, γ_{004}	0.649	1.914	0.657	0.597	1.817	0.645
Class Knowledge of Cognition, γ_{005}	0.118	1.125	0.529	0.164	1.178	0.541
Class Regulation of Cognition, γ_{006}	0.349	1.418	0.586	0.307	1.359	0.576
Class Pretest, γ_{007}	4.676	107.340	0.991	4.423	83.346	0.988
Treatment, γ_{008}	-0.051	0.950	0.487	-0.055	0.946	0.486
Discrimination Slope, γ_{100}	-1.215	0.297	0.229	-1.067	0.344	0.256
Pretest Slope, γ_{110}	3.379	29.341	0.967	3.100	22.198	0.957
Difficulty Slope, γ_{200}	0.814	2.257	0.693	0.730	2.075	0.675
Class Enjoyment, γ_{201}	-1.384	0.251	0.200	-1.319	0.267	0.211
Class Motivation, γ_{202}	0.338	1.402	0.584	0.290	1.336	0.572
Class Self Confidence, γ_{203}	0.961	2.614	0.723	0.894	2.445	0.710
Class Value, γ_{204}	0.118	1.125	0.529	0.158	1.171	0.539
Class Knowledge of Cognition, γ_{205}	0.138	1.148	0.534	0.090	1.094	0.522
Class Regulation of Cognition, γ_{206}	0.35	1.419	0.587	0.363	1.438	0.590
Class Pretest, γ_{207}	-0.536	0.585	0.369	-0.257	0.773	0.436
Treatment, γ_{208}	-0.231	0.794	0.443	-0.208	0.812	0.448
Motivation Slope, γ_{210}	-0.199	0.820	0.450	-0.187	0.829	0.453
Self Confidence Slope, γ_{220}	0.162	1.176	0.540	0.154	1.166	0.538
Pretest Slope, γ_{230}	-0.792	0.453	0.312	-0.583	0.558	0.358
Complexity Slope, γ_{300}	0.08	1.083	0.520	0.097	1.102	0.524

Table 104 (Continued)
Predicted Values for Final Class Model

	Unit Specific Model			Population Average Model		
Combined Effects	$\hat{\eta}$	Odds	Probability	$\hat{\eta}$	Odds	Probability
Int + Mean Discrimination + Mean Difficulty + Low Complexity + Mean Class Enjoyment, Self Confidence, Motivation + Class Mean Pretest + Std Mean Pretest + Std Mean Motivation + Control	-0.281	0.755	0.430	-0.274	0.760	0.432
Int + Mean Discrimination + Mean Difficulty + Low Complexity + Mean Self Confidence, Motivation + Class Mean Pretest + Std Mean Pretest + Std Mean Motivation + 1SD Above Mean Class Enjoyment + Control	-0.762	0.467	0.318	-0.740	0.477	0.323
Int + Mean Discrimination + Mean Difficulty + Low Complexity + 1SD Above Class Mean Self Confidence, Enjoyment, Motivation + Class Mean Pretest + Std Mean Pretest + Std Mean Motivation + Control	-0.201	0.818	0.450	-0.200	0.819	0.450
Int + Mean Discrimination + Mean Difficulty + Low Complexity + Class Mean Self Confidence, Enjoyment, Motivation, Pretest + Std Mean Pretest, Motivation + Treatment	-0.401	0.670	0.401	-0.382	0.682	0.406

The computation of probabilities from the fixed effects in Table 104 proceeded as in the item and student models, using Equations 33 and 34. The combined effects, however,required a consideration of the effects of class variables on student variables before combining student effects with item effects to produce the predicted value.

The first combined effect predicted the probability of a misconception error on an item of mean discrimination, mean difficulty, and low complexity for a student with mean pretest misconceptions and motivation in a control class with mean self confidence, enjoyment and motivation. Because all student and class level variables were centered (group and grand centered respectively), the student and class mean values produced a zero effect on the predicted probability. The intercept logit for this combined effect was 0.268. The discrimination impact was -1.215 (the coefficient logit) • 0.8 (the average discrimination) = -0.972, the value added to the intercept logit. The difficulty impact was 0.814, the coefficient logit • 0.52, 1 SD above average difficulty = 0.4233, the value added to the intercept logit. Complexity was coded as a dichotomous variable in which low complexity was coded as 0. Therefore, the predicted logit, $\hat{\eta}$, for this situation was 0.268 + -0.972 + 0.4233 = -0.281. The associated probability of misconception error (using Equations 33 and 34) was 0.430 in the sample. The same computational process was used for the population average model and subsequent combined effect examples.

The second and third combined effects from Table 104 represent the probability of a misconception error on a hard item (1 SD above mean Difficulty = 0.52) of mean discrimination (0.8) and low complexity (see Table 97 for difficulty and discrimination values). For combined effect 2, a student who had an average pretest misconception score in a control class with average pretest misconception, class enjoyment, motivation, value, and self confidence scores was predicted by the model to have a probability of a misconception error of 0.430 in the sample and 0.432 in the population. For the same student (Combined Effect 3), if the class enjoyment level increased by one standard deviation (0.3 units), the probability of a misconception error reduced to 0.318 in the sample and 0.323 in the population (Table 104).

The fourth combined effect examines the relationship of self confidence with misconception error probabilities. A student with pretest misconception score equal to the mean, on a hard item of average discrimination and low complexity, in a control class with average enjoyment, motivation, value, and pretest mean misconceptions but one standard deviation above the mean for self confidence had a misconception error probability of 0.450 for both the sample and population (Table 104).

The treatment condition had an indirect effect on the probability of a misconception error by impacting the item difficulty slope. A student with a mean pretest misconception score and motivation in a treatment class with average enjoyment, self confidence, value, and pretest misconception was predicted to have a probability of misconception error of 0.401 in the sample and 0.406 in the population on an item of average discrimination and difficulty and low complexity (Table 104).

Summary of Results

Three analyses were conducted to examine the nature of mathematical misconceptions. Document analysis of student responses was used to distinguish between errors due to factors other than misconceptions and errors representing misconceptions. The coding from this analysis was used for the subsequent quantitative analyses.

The quantitative analyses included two separate investigations. First, the relationship of misconceptions in each content area was examined using structural equation modeling. Six models were compared, all of which returned high goodness of fit indices and parsimony indices greater than 0.5. All models were calibrated using modification indices to reduce the chi-squared value. The final model for all six hypotheses validated well across a randomly chosen sample.

Second, the impact of item, student, and class characteristics on misconception errors was investigated through a three level hierarchical generalized linear model. In this model, the item and student variance was statistically non-significant, but the between class variance was significant. Only class knowledge of cognition was a significant predictor of misconception errors on a particular task. Based on these analyses, the item and student levels were combined into a single student level, in which the new outcome variable was the percentage of misconception errors for each student on the posttest. In this new model, student and class variance was statistically significant. Student mathematics self confidence and student percentage of misconceptions on the pretest were significant predictors of posttest misconception percentages. These two student variables explained 29% of student variance. Class enjoyment of mathematics, value of mathematics, and pretest misconception percentage were statistically significant class predictors of student posttest misconception percentage. The treatment condition of the class was not a statistically significant predictor. The removal of non-significant class predictor variables resulted in a significant amount of unexplained variance being returned to the model, so interaction effects were added to the model. The interaction of class value of mathematics and class knowledge of cognition was a statistically significant predictor of student posttest misconceptions.

These analyses offer information about the nature of misconceptions in mathematics that may lead to better assessment of misconceptions and interventions to address misconceptions. The following chapter discusses the implications of these results.

CHAPTER 5

DISCUSSION

This chapter presents a discussion of the results provided Chapter 4. Three analyses were conducted to investigate the nature of misconceptions in mathematics: (1) analysis of student response patterns on the mathematics knowledge test (NAEP items); (2) comparison of hypothesized structural models representing the relationships among content area misconceptions; and, (3) examination of the impact of item, student, and class characteristics on misconception errors. Item characteristics were measured using Item Response Theory on the NAEP mathematics knowledge test. Student characteristics included attitudes toward mathematics (ATMI Enjoyment, Motivation, Self Confidence, and Value scales), metacognitive awareness (MAI Knowledge of Cognition and Regulation of Cognition scales), and pre- and post-test misconception and percent correct scores. Class characteristics consisted of aggregated scores for each student characteristic along with indicator variables for treatment condition and type of mathematics class. Through these three analyses, seven key findings emerged.

1. Content area is not the most effective way to classify mathematics misconceptions; instead, five underlying misconceptions affect all four content areas.
2. Mathematics misconception errors often appear as procedural errors.
3. A classroom environment that fosters enjoyment of mathematics and value of mathematics are associated with reduced misconception errors.

4. Higher mathematics self confidence and motivation to learn mathematics is associated with reduced misconception errors.
5. Probability misconceptions do not have a causal effect on rational numbers, algebra, or geometry misconceptions.
6. Rational number misconceptions do not have a causal effect on probability, algebra, or geometry misconceptions.
7. Probability instruction may not affect misconceptions directly, but it may help students develop skills needed to bypass misconceptions when solving difficult problems.

Analysis 1: Misconception Error Analysis

Two key findings. The first analysis of the present study presented patterns of student responses on the mathematics knowledge test composed of NAEP items. This document analysis validated hypotheses about how misconceptions would result in error choices for eight of the 17 items. Misconception error choices for the other nine items were adjusted to align with observed student responses (as shown in Table 18) before proceeding with the quantitative analyses. The observed patterns of misconception errors revealed an important aspect of mathematics misconceptions, Key Finding 1: On a wide array of mathematical problems, a very small number of fundamental misconceptions (five) appeared to account for a large proportion of the observed errors (70.49%). All five of these core misconceptions (i.e., Absolute/Relative Comparison, Additive/Multiplicative Structure, Spatial Reasoning, Variable Meaning, and Rational Number Meaning Misconceptions) appeared in multiple mathematics content areas.

Another conclusion emerged from the analysis of student response patterns, Key Finding 2: Misconception error explanations relied on procedural knowledge isolated from conceptual knowledge (as described in Figure 10). Previous studies (e.g., Agnoli & Krantz, 1989; Kahneman & Tversky, 1972; Falk, 1992) have also indicated that reliance on judgmental heuristics may be an important factor in the development of mathematics misconceptions.

> Our task as mathematics educators is to distinguish between those circumstances in which judgmental heuristics can adversely affect stochastic thinking and those in which the heurisitcs are useful and desirable. And we are obliged to point out the differences to our students. It is not that there is "something wrong" with the way our students think. It is just that they (and we) tend to carry useful heuristics beyond their relevant domain (Shaughnessy & Bergman, 1993, p. 184).

The analysis of the present study indicates that connecting procedural knowledge to conceptual knowledge may help teachers and students make these distinctions. Hiebert and Grouws (2007) described two observable features for a classroom that focuses on developing conceptual understanding: (1) Teaching focuses explicitly to connections between facts, procedures, and ideas, and (2) Students are allowed to struggle with important mathematical concepts. Development of these two features in a classroom may help teachers identify the reasoning behind errors that emerge from misconceptions.

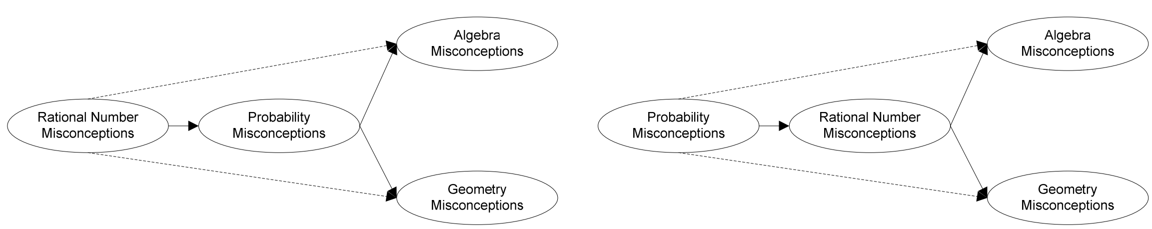
Detecting mathematics misconceptions. NAEP released items were compiled "as is," without any changes for the mathematics knowledge test (Appendix M). By doing so, the NAEP-established item content and concurrent criterion validity could be transferred to the present study (Daro et al., 2007). The compiled instrument also exhibited acceptable internal consistency and test-retest reliability. Despite these qualities, the instrument failed to adequately differentiate between misconceptions.

The ambiguity in student explanations for several items indicated that the validity of the items did not necessarily extend to measuring misconceptions. For example, the question of whether content area or type of underlying misconception category is a better way to organize mathematical misconceptions cannot be answered by the present study — some item responses indicated multiple types of misconceptions (e.g., Item 7, Choice D; Item 17, Choice C). Such a question might be answerable using a multi-trait, multi-method structural equation model, but a model of this type, based on the present instrument would most likely require several cross-loadings that would make the model structurally unstable (i.e., no amount of iterations can yield a solution), such as was seen in Models A, B, and C of the structural analysis. Therefore, I recommend that such a study begin by altering the present instrument to focus directly on observed misconception responses. For example, in Item 17, the uniformity heuristic sometimes represented an absolute/relative comparison misconception, an additive/multiplicative structure misconception, a rational number meaning misconception, or a combination of these misconception types. To distinguish misconceptions more readily, it may be necessary to include explanations with possible answers. For example, instead of simply offering the choice "three," a revised item might offer "one because the numerator is one (absolute/relative comparison misconception), "three because the denominator is three" (rational number meaning misconception), and "three because R and S have equal faces" (additive/multiplicative structure misconception via the uniformity heuristic). Without such differentiation, misconception content validity for closed-response items will be difficult to establish. A study to develop and validate such a misconception instrument may be a necessary first step to replicating and advancing the present investigation.

Analysis 2: Content Area Misconceptions

Two key findings. The second analysis of the present study compared six hypothesized relationships between misconceptions in probability, rational numbers, algebra, and geometry as shown in Figure 1. Multiple studies found during the literature review (e.g., Agnoli & Krantz, 1989; Falk, 1992; Freudenthal, 1970, 1973, 1983; Schield, 2006; Shaughnessy & Bergman, 1993; Warren, 2000; Watson & Shaughnessy, 2004) suggested that misconceptions in probability and rational numbers may hold a causal predictive position relative to those of algebra and geometry. They did not, however, suggest which might be the primary causal factor or if both acted together as causal indicators. The results of the present study suggested Key Findings 5 and 6: Probability misconceptions do not have a causal effect on rational numbers, algebra, or geometry misconceptions; Rational number misconceptions do not have a causal effect on probability, algebra, or geometry misconceptions. These results also reinforced Key Finding 1: Content area is not the most effective way to classify mathematics misconceptions.Interpretation of this analysis focused on issues of model stability and comparisons of the final models using the goodness of fit indices.

Model stability. Model C (Figure 24) exhibited instability throughout the structural model analysis. Instability refers to the iterative process of SEM being unable to determine a best fitting solution. In the case of Model C, Lisrel 8.72 identified the fitted covariance matrix as "not positive definite," meaning that the determinant of the solution matrix was either less than or equal to zero. This error indicates an unstable model for two reasons. (1) If the determinant of the fitted matrix equals zero, then the matrix is not invertible: The minimization function requires that the fitted matrix be inverted to find a solution. (2) A negative determinant allows the matrix to be inverted mathematically, but the determinant is negative in a symmetric matrix only when elements on the diagonal are negative (Wothke, 1993). These diagonal elements represent variance (Byrne, 1998), so negative values present serious problems for interpretation and fitting the estimated matrix to the observed matrix. Wothke (1993) described two reasons for non-positive definite matrices, collinearity and overparameterization. In both cases, removal of unnecessary or redundant parameters can allow the analysis to proceed. In the case of Model C, both rational number and probability misconceptions could not predict both algebra and geometry misconceptions simulateously without creating a non-positive definite fitted covariance matrix. The structural portion of Model C was therefore adjusted to discover the source of the error, which turned out to be the crossloadings from both independent variables (rational number and probability) to both dependent variables (algebra and geometry). Removal of either the crossloadings or direct effects eliminated the non-positive definite matrix problem. I proceeded to ask whether non-positive definite matrices would have occurred across imputations in Models A and B if the models had specified additional structural parameters. The dotted line regression weights in Figure 32 represent each additional parameter added to the models.



A **B**

Figure 32. Post Hoc Hypothesized Structural Parameters for Models A and B (Figure 24)

By adding the dotted line regression pathways in Figure 33 one at a time and together, both models resulted in non-positive definite matrices that prevented the analysis from computing an admissible solution. The consistent pattern of non-convergence across three different models led to the conclusion that the addition of too many structural parameters created unstable models when both rational number and probability misconceptions act as predictors of both algebra and geometry misconceptions. One interpretation of this model behavior may be that rational number misconceptions and probability misconception may impact algebra and geometry misconceptions through the other or directly, but they do not appear to act as both direct causes and moderators simultaneously.

Goodness of fit statistics. Each competing model in the calibration sample demonstrated excellent goodness of fit statistics with the exception of the GFI, and the GFI statistic was consistently moderate across all models. The dichotomous nature of the data meant that non-normality necessarily existed in the measurements, which has been reported to influence the value of GFI (Hu & Bentler, 1995). None of the models demonstrated superiority over the others (see Table 84).

Conclusions. Based on these analyses, none of the models represented the relationship between misconceptions across content areas better than the others. Key Finding 5 and 6 emerged from these results: Probability misconceptions do not have a causal effect on rational numbers, algebra, or geometry misconceptions; and, Rational number misconceptions do not have a causal effect on probability, algebra, or geometry misconceptions.

Multicollinearity between content area misconceptions may account for the lack of causal relationships. If multicollinearity were present, it could have caused the different causal models to be statistically indistinguishable. To test for multicollinearity between content area misconceptions, pooled Pearson correlations between factor scores were computed across the five imputed data sets (Table 105).

Table 105
Pooled Intercorrelations Between Content Area Misconception Scores

Subscale	1	2	3	4
1. Algebra	—	$.31^{**}$	$.36^{**}$	$.35^{**}$
2. Geometry		—	$.27^{**}$	$.23^{**}$
3. Rational Number			—	$.32^{**}$
4. Probability				—

Note: N = 1133; ** $p < 0.001$

All correlation coefficients in Table 105 were statistically significant, ranging from 0.23 to 0.37. These values mean that 5% to 13% of the variance between any two variables can be accounted for by multicollinearity. Taken together, these correlations confirmed that mutlicollinearity was a significant factor in the model analysis.

By considering the qualitative analysis of student response patterns along with the present structural model analysis, the source of the multicollinearity can be traced. In the qualitative analysis, students sometimes responded to a misconception item because of different misconceptions (e.g., Item 4, Response B and Item 1, Response E). In other items, students chose one distractor because of one type of misconception, but chose another response because of a different misconception (e.g., Item 17). Because of this lack of discrimination within some misconception responses, examining a theoretical model of the underlying misconceptions (i.e., the meaning of rational numbers, additive/multiplicative structures, spatial reasoning, absolute/relative comparison, and the meaning of variables) was not possible with any degree of reliability. The results of the present study may strengthen the notion that there exists a core set of misconceptions that span content areas; indeed, the lack of difference among content area based models indicates that these underlying misconception structures that influence reasoning in all four content areas may be more important than the content area to understanding how mathematical misconceptions develop and how they can be better addressed.

To test this hypothesis, a second order factor was added to Model F (Figure 31). I hypothesized two potential outcomes for this new model. If the model demonstrated significant improvement over the other models, then this first outcome would mean that content area factors may model misconceptions well, but a higher order, fundamental mathematics misconceptions would account for the linearity between them. If, on the other hand, the new model did not demonstrate significant differences with the other models, then content area factors may not be the best way to model mathematical misconceptions.

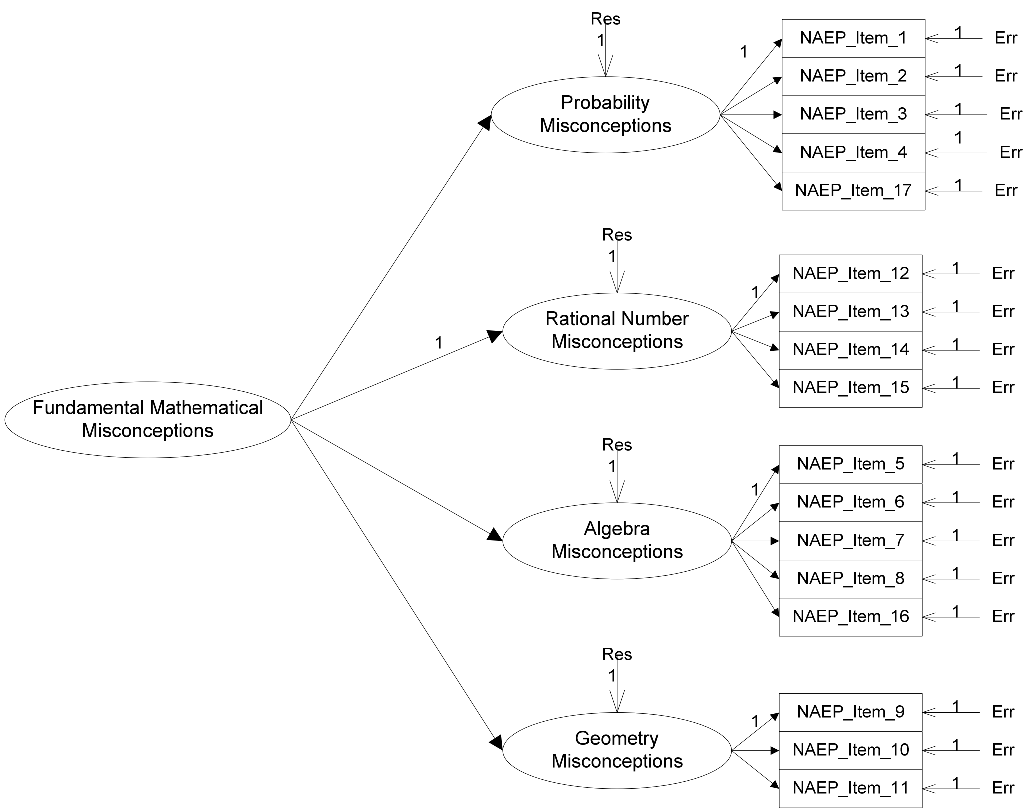


Figure 33. Possible Second Order Factor Model to Explain Content Area Misconception Multicollinearity.

The addition of this second order factor required checking that the new level was over-identified. With 10 data moments, 3 regression weights, 1 variance, and 4 residuals, the second order construct was overidentified with 10 – (3+1+4) = 2 degrees of freedom. Overall, the model had 115 degrees of freedom.

The resulting goodness of fit statistics indicated that this model behaved no differently than Models A-F (Table 106).

Table 106
Goodness of Fit Indices for Second Order Post Hoc Hypothesized Model

χ^2	CFI	GFI	RMSEA	RMSEA 90% CI	SRMR	PGFI	PCFI	ECVI
604.17	0.99	0.87	0.024	[0.009, 0.035]	0.075	0.66	0.751	1.41

This structural comparison supported the hypothesis that the multicollinearity between content area misconceptions cannot be explained by the addition of a single factor. Previous studies (e.g., Agnoli, 1987; Agnoli & Krantz, 1989; Battista, 2007; Clements & Battista, 1992; De Bock et al., 2002; Falk, 1992; Kahneman & Tversky, 1973a, 1973b, 1982; Küchemann, 1978; Lamon, 1999; Shaughnessy & Bergman, 1993; Van Dooren et al., 2003; Warren, 2000; Watson & Shaughnessy, 2004) have indicated at least five potential cross-content misconceptions that may explain the multicollinearity found in the present study, reinforcing Key Finding 1. Figure 34 portrays a possible structure for these underlying misconceptions and how they affect misconceptions in each content area.

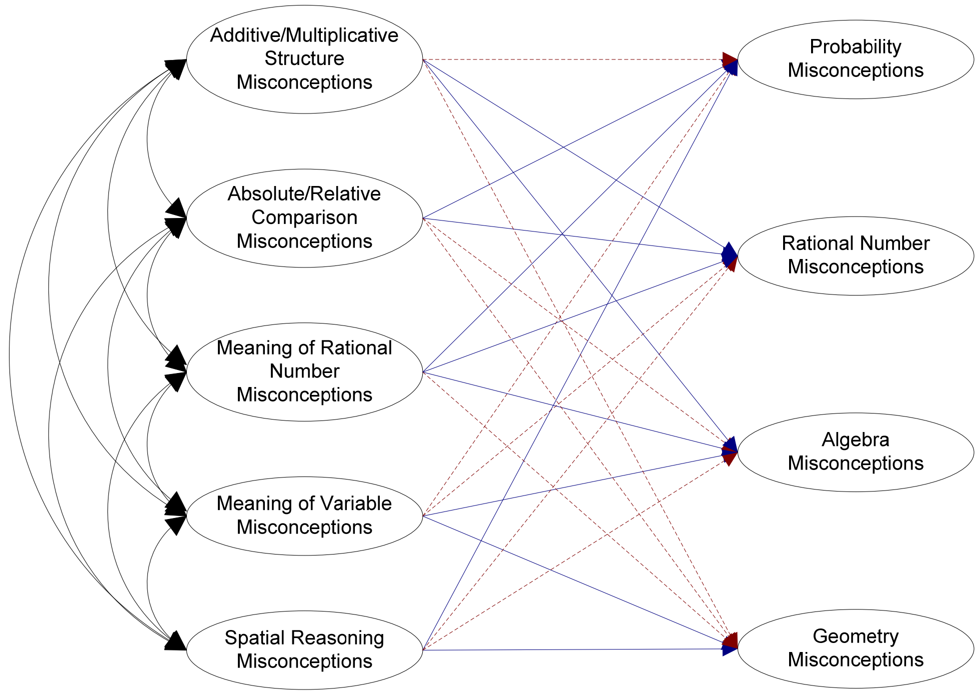


Figure 34. Hypothetical Structure for Underlying Mathematical Misconceptions.

The solid-line regression weights in Figure 34 represent relationships observed in the qualitative analysis of student responses. The dotted-line regression weights represent potential relationships between underlying and content area misconceptions that were not observed in the qualitative analysis. Whether their absence was due to the lack of a relationship or simply an artifact of the assessment instrument is not entirely clear — the dotted lines are not necessarily weaker relationships. For example, the meaning of rational number misconceptions very likely impacts geometry misconceptions relating to similarity concepts, but no items in the assessment instrument measured similarity concepts.

Additionally, the relationship between the underlying misconceptions is not clear at this time. There may well be a causal structure between these factors; some may also be completely uncorrelated with others. Future studies may wish to measure these relationships.

Implications for curriculum development. The NCTM (2000) *Principles and Standards for School Mathematics* offered a conceptual framework to organize the emphasis of each content strand from Pre-K to Grade 12 (Figure 9). The structural analysis of the present study suggests that such an organizational structure may be unable to conceptualize fundamental concepts to learning mathematics that often result in misconceptions when ignored. A modified scheme adds an extra dimension to the NCTM framework (Figure 35).

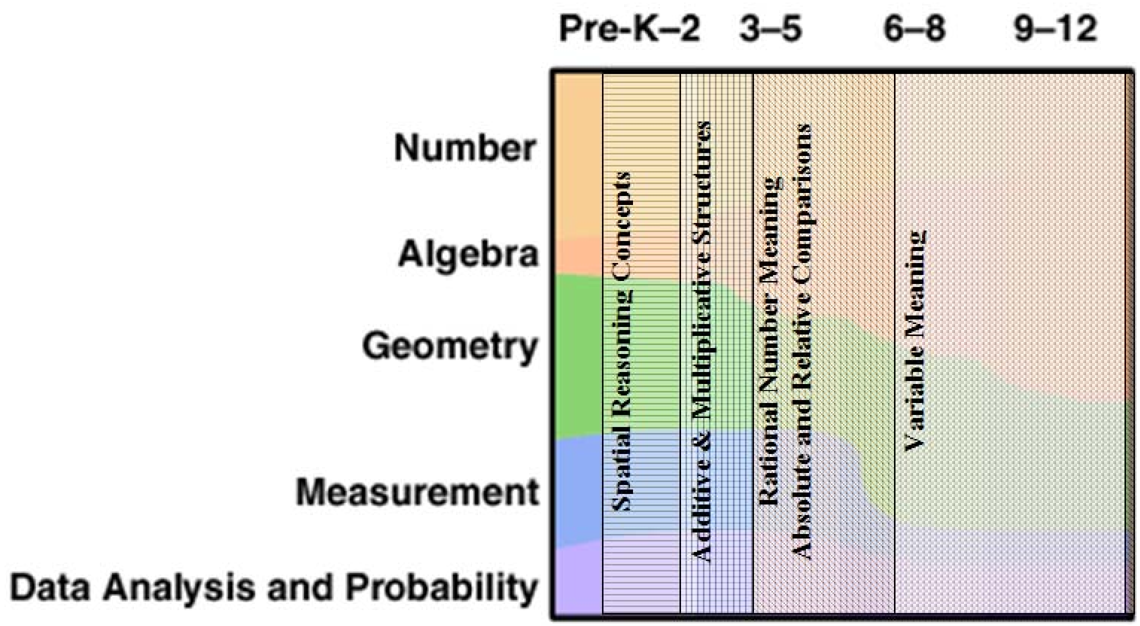


Figure 35. NCTM (2000) Modified Content Emphases Including Fundamental Mathematics Concepts

Figure 35 depicts these fundamental mathematics concepts as progressive stages of learning throughout a child's education. As a child progresses through grade school, the learning of these concepts can follow two paths. (1) If left unchecked, these concepts develop into misunderstandings about ideas, and misconceptions may develop. These misconceptions may compound as new learning barriers are encountered. When learning is focused primarily on developing procedural knowledge, the resultant rules, heuristics, and formulas developed in isolation from conceptual knowledge lead to instrumental understanding and misunderstanding about the meaning of important mathematical ideas (as shown in Figure 10). (2) If misconceptions are addressed through an intervention that reinforces the meaning of mathematics ideas and the connections between ideas, then the learning barriers in Figure 35 may bolster rather than hinder the development of student understanding (Resnick, 1983).

Textbooks may also improve their effect on student mathematics learning by integrating the barriers of Figure 35 throughout lesson sequences. Consider the example of a linear function definition presented by an algebra textbook. In Chapter 1, this problem was described as focusing primarily on a prescription for recognizing linear functions. The textbook description also connected the meaning of lines with the shape of the graph (i.e., spatial reasoning). The textbook did not, however, address the meaning of the variables x, y, or $f(x)$. It also failed to address the meaning of the quantities m and b. Furthermore, no comparison was made of the similarities and differences in m and b from x and y. A discussion of the meaning of m can be used to address the differences between additive and multiplicative structures as described by Warren (2000).

The potential changes to mathematics curricula supported by the present study may add a layer of complexity to the way mathematics content is organized. This complexity may have a direct impact on the ability of educators to provide materials to help address the barriers students encounter when learning mathematics.

Analysis 3: Factors Influencing Misconception Errors

Three key findings. The final analysis of the present study examined the impact of item, student, and class characteristics on misconception errors. The results of this analysis led to Key Findings 3, 4, and 7: A classroom environment that fosters enjoyment of mathematics or value of mathematics helps reduce student misconception errors; Higher mathematics self confidence reduces misconception errors; and, Probability instruction may not affect misconceptions directly, but it may help students develop skills needed to bypass misconceptions when solving difficult problems.

The two level model revealed significant predictors of misconceptions for both student and class characteristics. Student mathematics self confidence (ATMI self confidence scale) and pretest misconception error percentages (NAEP instrument) accounted for 29% of the student variance in posttest misconception error percentages (NAEP instrument). Class enjoyment of mathematics (ATMI enjoyment scale) and the class value of mathematics (ATMI value scale) also had a statistically significant effect on posttest misconception errors. The between-class variance in the unconditional model (see Table 89) was 0.0073 ($p < 0.001$). The contextual model that included the statistically significant class variables (see Table 92) reduced the between class variance to 0.0005 ($p = 0.016$). This reduction represented a 93.15% reduction. The three level model also accounted for 98% of the class level variance. Such large reductions in variance indicates that a large percentage of class effects on misconceptions may lie in the factors measured by the ATMI scales. If true, then educators can begin focusing on improvement of these factors within a class to reduce misconceptions.

Implications for teaching mathematics. Traditional mathematics instruction has relied primarily on teacher-centered epistemologies (Stigler & Hiebert, 1999). This investigation began with the assumption that student-centered instructional approaches have a more positive effect on student mathematics learning than traditional, teacher-centered strategies. The present study supported this assumption and extended it to addressing misconceptions. Higher mathematics self confidence, value, and enjoyment were associated with a decline in misconceptions; the development of a positive learning environment may therefore be a critical component to helping students traverse the learning barriers in Figure 35.

Maher and Tetrault (2001) described four epistemological components critical to developing such a positive learning environment: mastery, voice, authority, and positionality. First, mastery involves struggle and engagement with a body of knowledge. Instead of merely absorbing information, students grapple with difficulties of understanding. Rather than the final product as end goal, mastery refers to the continual process of working and re-working information into knowledge. Second, the fashioning of one’s voice in mathematics means to bring one's personal experiences, questions, and perspectives to the mathematics being studied. Third, the concept of authority refers to the source of mathematical knowledge in a classroom. Maher and Tetrault (2001) and Shrewsbury (1993) described a climate of shared mathematical authority: Students and teachers share the knowledge and understanding of important mathematical ideas in such an environment. Authority refers to the relationship between students and teachers collectively with mathematical knowledge. Fourth, positionality refers to the relationships between an individual and mathematical knowledge along with the interactions of these within- and between-student relationships.

As teachers seek to help students turn the barriers of Figure 35 into opportunities to reinforce fundamental mathematics concepts, the development of a student-centered environment may be a foundational component for any strategy. Previous studies (e.g., Slavin & Karweit, 1982; Slavin & Lake, 2008; Slavin et al., 2009) have found that student-centered teaching approaches provide benefits to student achievement. If these environments are to offer the most benefit to avoiding and addressing misconceptions, then students must be given opportunities to struggle with important mathematical ideas and their connections (Hiebert & Grouws, 2007; Kieran, 1989, 1992, 2007).

An ontological perspective from the present study also offers insight for helping students overcome the learning barriers in Figure 35. The examination of student misconception error explanations revealed a consistent pattern: Misconception errors occurred when students relied on procedures isolated from meaning and mathematical structure. This pattern suggests that mathematics is best understood as an organized structure of meanings and connections rather than procedures.

> Teachers should strive to organize the mathematics so that fundamental ideas form an integrated whole. Big ideas encountered in a variety of contexts should be established carefully, with important elements such as terminology, definitions, notation, concepts, and skills emerging in the process (NCTM, 2000, p. 14)

In combination with the epistemological implications described above, the present study found that student-centered, concept-focused mathematics classrooms may be the most effective learning environment for turning fundamental mathematics barriers into opportunities to learn.

Final Thoughts: Pedagogy and Mathematics Misconceptions

Traditional mathematics pedagogy may be even more detrimental to student learning than described by Welch (1978), Stigler and Hiebert (1997), and Manoucheri and Goodman (2001). In addition to losing the opportunity to struggle with important mathematics, traditional pedagogy also removes students from a position in which they can value or enjoy mathematics, individually or collectively (Shrewsbury, 1993). Traditional pedagogy positions students to receive and react to goals set out by the teacher rather than allowing them part of the leadership process that helps them develop metacognitive knowledge and skills (Maher & Tetrault, 2001; Shrewsbury, 1993). Conceptual instruction, on the other hand, guides students to developing their own understanding of the meaning of important mathematical ideas and the connections between these ideas through authentic intellectual struggle (Rousseau, 1976; Stone, Alfeld, & Pearson, 2008). The findings of the present study suggest that conceptually focused instruction may position students to grapple with complex, abstract mathematical in a way that helps develop relational understanding of relevant mathematical structures, thereby reducing misconceptions or turning misconceptions into learning experiences.

When targeting misconceptions, attending to the underlying structure of misconceptions that appear in all four content areas may be more effective than targeting the observable errors resulting from those misconceptions. Future investigations of mathematics misconceptions may best begin by developing a more refined instruments for identifying and categorizing misconceptions and potential causal structures

The present study explored the use of probability instruction as an intervention to target fundamental mathematics misconceptions. The treatment condition had a statistically significant impact on the effect of item difficulty on misconception error probabilities, and several other important statistically significant factors were also identified. None of the hypothesized structures of content area misconceptions could be identified as a better fit to the data. This outcome led to the discovery of a high degree of multicollinearity between content area misconceptions, which supported the notion of an underlying mathematics misconception structure. Preliminary analysis of a second-level underlying structure showed promise for this approach to understanding mathematical misconeptions. This finding offers a radically different perspective on the nature of mathematics and mathematics learning. Furthermore, only five foundational concepts appeared to be fundamental to learning mathematics. Attending to these five foundational concepts may allow mathematics teaching in any single area to fundamentally impact the learning of other mathematics areas. Ignoring this small set of foundational concepts may allow the development of a formidable obstruction at a level that can inhibit and perhaps derail the mathematics future of students. Such an astounding notion may indicate that finding ways to identify and address these foundational concepts and their connections to a particular mathematics area should be one of the primary, critical tasks for mathematics educators.

REFERENCES

Adamson, J. W. (Ed.). (1922). *The educational writings of John Locke*. Cambridge: University Press.

Agnoli, F. (1987). Teaching implications of misconceptions in probability and statistics. In J. Novak (Ed.), *Proceedings of the second international seminar: Misconceptions and educational strategies in science and mathematics* (Vol. I, pp. 1-5). Ithaca, NY: Cornell University.

Agnoli, F., & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. *Cognitive Psychology, 21,* 515-550.

Ahadi, S. A., & Rothbart, M. K. (1994). Temperament, development, and the Big Five. In C. F. Halverson, G. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 189-207). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ainley, J., Bills, L., & Wilson, K. (2004, July). *Constructing meanings and utilities within algebraic tasks.* Paper presented at the 28th conference of the International Group for the Psychology of Mathematics Education, Bergen, Norway.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second International Symposium on Information Theory* (pp. 267-281). Budapest: Akademiai Kiado.

Akaike, H. (1983). Information measures and model selection. *Bulletin of the International Statistical Institute: Proceedings of the 44th Session, 1*, 277-290.

Alomar, B. O. (2007). Personal and family factors as predictors of pupils' mathematics achievement. *Psychological Reports, 101,* 259-269.

Alpine Media Corporation. (2003). *The Temperament Sorter II: A statistical study*. Orem, UT: Author.

Amenkhienan, C. A., & Kogan, L. R. (2004). Engineering students' perceptions of academic activities and support services: Factors that influence their academic performance. *College Student Journal, 38,* 523-540.

Amir, R., Frankl, D. R., & Tamir, P. (1987). Justifications of answers to multiple choice items as a means for identifying misconceptions. In J. Novak (Ed.), *Proceedings of the second international seminar: Misconceptions and educational strategies in science and mathematics* (Vol. I, pp. 15-25). Ithaca, NY: Cornell University.

Anderson, S. E. (1990). Worldmath curriculum: Fighting Eurocentrism in mathematics. *Journal of Negro Education, 59*, 348-359.

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice, 48,* 12-19.

Ashman, S. M. (2001). Consumer choice models with customer loyalty programs in retail food stores. *Dissertation Abstracts International – A, 61*(11), 4484. (UMI No. AAT 9994493)

Baker, F. B., & Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques*. New York: Marcel Dekker.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*, 389-409.

Bandalos, D. L. (1993). Factors influencing cross-validation of confirmatory factor analysis models. *Multivariate Behavioral Research, 28,* 351-374.

Bandalos, D. L. (2009, April). *A comparison of fit index behavior, parameter estimate, and standard error bias for some mean- and variance-adjusted estimators.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Bar-Hillel, M., & Falk, R. (1982). Some teasers concerning conditional probabilities. *Cognition, 11,* 109-122.

Barkatsas, A., Kasimatis, K., & Gialamas, V. (2009). Learning secondary mathematics with technology: Exploring the complex interrelationship between students' attitudes, engagement, gender and achievement. *Computers & Education, 52,* 562-570.

Baroudi, Z. (2006). Easing students' transition to algebra. *Australian Mathematics Teacher, 62*, 28-33.

Barry, R. R. (1988). The application of sabermetrics to the teaching and coaching of collegiate baseball. (Doctoral dissertation, Middle Tennessee State University, 1988). *Dissertation Abstracts International-A*, *50*(02), 384.

Battista, M. T. (2007). The development of geometric and spatial thinking. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 843-908). Reston, VA: National Council of Teachers of Mathematics.

Baturo, A. R. (1994, August). *A cognitive analysis of misconceptions in year 12 students' understanding of elementary probabilistic notions*. Paper presented at the 3rd International Seminar of Misconceptions and Educational Strategies in Science and Mathematics, Ithaca, NY.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John

Canton, A. M. F. R. S. *Philosophical Transactions (1683-1775), 53,* 370-418.

Behr, M. (1980). How children view the equal sign. *Mathematics Teaching, 92,* 13-15.

Behr, M. J., Harel, G., Post, T. R., & Lesh, R. (1992). Rational number, ratio, and proportion. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 296-333). Reston, VA: National Council of Teachers of Mathematics.

Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. *American Journal of Physics, 62,* 750-762.

Bell, A. (1987). Misconceptions, conflict and discussion in the teaching of graphical interpretations. In J. Novak (Ed.), *Proceedings of the second international seminar: Misconceptions and educational strategies in science and mathematics* (Vol. I, pp. 46-58). Ithaca, NY: Cornell University.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.

Berg, K. F. (1993, April). *Structured cooperative learning and achievement in a high school mathematics class*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Blanco, L., & Garrote, M. (2007). Difficulties in learning inequalities in students of the first year of pre-university education in Spain. *Eurasia Journal of Mathematics, Science & Technology Education, 3,* 221-229.

Bloodbook.com. (2007). *Blood type facts.* Retrieved on May 14, 2009, from http://www.bloodbook.com/type-facts.html

Boland, P. J., & Nicholson, J. (1996). The statistics and probability curriculum at the secondary school level in the USA, Ireland and the UK. *The Statistician, 45,* 437-

446.

Bong, M. (2008). Effects of parent-child relationships and classroom goal structures on motivation, help-seeking avoidance, and cheating. *Journal of Experimental Education, 76,* 191-217.

Booth, L. R. (1986). Difficulties in algebra. *Australian Mathematics Teacher, 42*, 2-4.

Boston, M. D., & Smith, M. S. (2009). Transforming secondary mathematics teaching: increasing the cognitive demands of instructional tasks used in teachers' classrooms. *Journal for Research in Mathematics Education, 40*, 119-156.

Bowles, S., & Levin, H. M. (1968). The determinants of scholastic achievement — An appraisal of some recent evidence. *The Journal of Human Resources, 3*, 3-24.

Brandt, A. E., & Pietras, C. J. (2008). Gambling on a simulated slot machine under conditions of repeated play. *The Psychological Record, 58,* 405-426.

Brick, J. M., Jones, M. E., Kalton, G., & Valliant, R. (2005). Variance estimation with hot deck imputation: A simulation study of three methods. *Survey Methodology, 31*, 151-159.

Briggs, J., Demana, F., & Osborne, A. R. (1986). Moving into algebra: Developing the concepts of variable and function. *Australian Mathematics Teacher, 42,* 5-8.

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research, 24,* 445-455.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136-162). Newbury Park, CA: Sage.

Brownell, W. A. (1935). Psychological considerations in the learning and teaching of arithmetic. In W. D. Reeve (Ed.), *The teaching of arithmetic: Tenth yearbook of* *the National Council of Teachers of Mathematics* (pp. 1-31). New York: Teachers College, Columbia University.

Brownell, W. A. (1956). Meaning and skill: Maintaining the balance. *Arithmetic Teacher, 34,* 18-25.

Bryk, A. S., & Raudenbush, S. W. (1988). Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. *Psychological Bulletin, 104,* 396-404.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming.* Mahwah, NJ: Lawrence Erlbaum Associates.

Byrne, B. M. (2009). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Capaldi, D. M., & Rothbart, M. K. (1992). Development and validation of an early adolescent temperament measure. *Journal of Early Adolescence, 12*, 153-173.

Capraro, M. M., Kulm, G., & Capraro, R. M. (2005). Middle grades: Misconceptions in statistical thinking. *School Science and Mathematics, 105,* 165-174.

Carraher, D. W., & Schliemann, A. D. (2007). Early algebra. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 669-706). Reston, VA: National Council of Teachers of Mathematics.

Carrell, S. E., Page, M. E., & West, J. E. (2009, May). *Sex and science: How professor gender perpetuates the gender gap (Working Paper No. 14959).* Cambridge, MA: National Bureau of Economic Research. (ERIC Document Reproduction Service No. ED505322)

Carter, J. A., Cuevas, G. J., Day, R., Malloy, C., Cummins, J., & Zike, D. (2010). *Geometry.* Columbus, OH: Glencoe/McGraw-Hill.

Carter, S. (2008). Disequilibrium and questioning in the primary classroom: Establishing routines that help students learn. *Teaching Children Mathematics, 15,* 134-137.

Castro, C. S. (1998). Teaching probability for conceptual change. *Educational Studies in Mathematics, 35,* 233.

Catanzaro, D. (1997). Course enrichment and the job charachteristics model. *Teaching of Psychology, 24*, 85-87.

Chambers, D. L. (1994). The right algebra for all. *Educational Leadership, 53,* 85-86.

Chang, Y. (2002, March). *Understanding and learning of function: Junior high school students in Taiwan.* Paper presented at the International Conference on Mathematics – "Understanding Proving and Proving to Understand," Taipei, Taiwan.

Chazan, D., & Yerushalmy, M. (2003). On appreciating the cognitive complexity of school algebra: Research on algebra learning and directions of curricular change. In J. Kilpatrick, W. G. Martin, & D. Schifter, *A research companion to "Principles and standards for school mathematics"* (pp. 123-135). Reston, VA: National Council of Teachers of Mathematics.

Clement, L. L. (2001). What do students really know about functions? *Mathematics Teacher, 94*, 745-748.

Clement, J. (1982). Algebra word problem solutions: Analysis of a common misconception. *Journal for Research in Mathematics Education, 13,* 16-30.

Clements, D. H., & Battista, M. T. (1992). Geometry and spatial reasoning. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp.

420-464). Reston, VA: National Council of Teachers of Mathematics.

Clotfelter, C. T., & Cook, P. J. (1991). Lotteries in the real world. *Journal of Risk and Uncertainty, 4,* 227-232.

Cobb, P., Gresalfi, M., & Hodge, L. L. (2009). An interpretive scheme for analyzing the identities that students develop in mathematics classrooms. *Journal for Research in Mathematics Education, 40,* 40-68.

Coes, L. (1993). Building fractal models with manipulatives. *Mathematics Teacher, 86,* 646-651.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (2nd ed.). Mahweh, NJ: Lawrence Erlbaum Associates.

Coleman, J. S. (1967a). '*Equality of educational opportunity,' reconsidered.* Washington, DC: U.S. Department of Health, Eduction & Welfare Office of Education. (ERIC Document Reproduction Service No. ED015893)

Coleman, J. S. (1967b). *The concept of 'Equality of educational opportunity.'* Washington, DC: U.S. Department of Health, Eduction & Welfare Office of Education. (ERIC Document Reproduction Service No. ED015157)

Coleman, J. S. (1968). *The evaluation of 'Equality of educational opportunity.'* Washington, DC: U.S. Department of Health, Education, & Welfare Office of Education. (ERIC Document Reproduction Service No. ED026721)

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity.* Washington DC: U.S. Department of Health, Eduction & Welfare Office of Education. (ERIC Document Reproduction Service No. ED012275)

CollegeBoard Advanced Placement. (2008). *Statistics course description.* New York:

CollegeBoard.com Inc.

Collis, K. F. (1975). *A study of concrete and formal operations in school mathematics: A Piagetian viewpoint*. Hawthorn, Victoria: Australian Council for Educational Research.

Confrey, J. (1987). 'Misconceptions' across subject matters: Science, mathematics and programming. In J. Novak (Ed.), *Proceedings of the second international seminar: Misconceptions and educational strategies in science and mathematics* (Vol. I, pp. 81-106). Ithaca, NY: Cornell University.

Confrey, J. (2006, November 11). Fuzzy policy, not 'Fuzzy math,' Is the problem. *Education Week*. Retrieved on January 19, 2009, from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=23096977&site=ehost-live

Confrey, J., & Lipton, A. (1985). Misconceptions research and the clinical interview. In S. K. Damarin & M. Shelton (Eds.), *Proceedings of the Seventh Annual Meeting of the North american Chapter of the International Group for the Psychology of Mathematics Education* (pp. 38-43), Columbus, OH.

Cooney, T. J., Davis, E. J., & Henderson, K. B. (1975). *Dynamics of teaching secondary school mathematics*. Boston: Houghton Mifflin.

Coppen, M. J. (1976). The development and evaluation of an individualized mastery programme in mathematics for low-ability secondary school students. *New Zealand Journal of Educational Studies, 11,* 132-142.

Crawford, M., Souders, J., & Pedrotti, L. (2009). *Cord algebra 1: Learning in context*. Waco, TX: Cord Communications.

Creswell, J. W. (2007). *Qualitative inquiry and research design* (2nd ed.). Thousand

Oaks, CA: Sage.

Crowley, M. L. (1987). The Van Hiele model of the development of geometric thought. In M. M. Lindquist & A. P. Shulte (Eds.), *Learning and teaching geometry, K-12: 1987 yearbook* (pp. 1-16). Reston, VA: National Council of Teachers of Mathematics.

Daggett, W. R. & Kruse, B. (1999). *Taming the educational dinosaur*. Rexford, NY: Leadership Press.

Daro, P., Stancavage, F., Ortega, M., DeStefano, L., & Linn, R. (2007). *Validity study of the NAEP mathematics assessment: Grades 4 and 8*. Washington, DC: American Institutes for Research. Retrieved March 28, 2010, from http://www.air.org/publications/documents/NAEP_Math_Validity_Study.pdf

Davis, B. (2007). Learners within contexts that learn. In T. Lamberg & L. R. Wiest (Eds.), *Proceedings of the 29th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* Stateline, NV: University of Nevada, Reno.

Davis, R. B. (1984). *Learning mathematics: The cognitive science approach to mathematics education*. Norwood, NJ: Ablex.

Davis, R. B. (1992). Reflections on where mathematics education now stands and on where it may be going. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 724-734). Reston, VA: National Council of Teachers of Mathematics.

Deal or No Deal. (2009). *Deal or no deal game show*: Play the TV game show online. New York: NBC.com. Retieved October 1, 2009, from http://www.nbc.com/Deal_or_No_Deal/game/flash.shtml

De Bock, D., Van Dooren, W., Verschaffel, L., & Janssens, D. (2002). Improper use of linear reasoning: An in-depth study of the nature and the irresistibility of secondary school students' errors. *Educational Studies in Mathematics, 50,* 311-334.

De Bock, D., Verschaffel, L., & Janssens, D. (1998). The predominance of the linear model in secondary school students' solutions of word problems involving length and area of similar plane figures. *Educational Studies in Mathematics, 35,* 65-83.

Dermitzaki, I., Leondari, A., & Goudas, M. (2009). Relations between young students' strategic behaviours, domain-specific self-concept, and performance in a problem-solving situation. *Learning and Instruction, 19,* 144-157.

Dietiker, L., Kysh, J., Hoey, B., & Salley, T. (2006). *Algebra connections* (Rev. ed.). Sacramento, CA: College Preparatory Mathematics.

Dietiker, L., Kysh, J., Hoey, B., & Sallee, T. (2007). *Geometry connections* (Version 3.1). Sacramento, CA: CPM Educational Program.

Douglas, J. B. (1970). On the teaching of probability and statistics at the pre-college level in Australia. In L. Råde (Ed.), *The teaching of probability and statistics* (pp. 303-318). Stockholm: Almqvist & Wiksell Förlag AB.

Edwards, T. G. (2000). Some 'Big ideas' of algebra in the middle grades. *Mathematics Teaching in the Middle School, 6,* 26-31.

Ellis, L. K., & Rothbart, M. K. (2001, April). *Revision of the Early Adolescent Temperament Questionnaire.* Poster session presented at the Biennial Meeting of the Society for Research in Child Development, Minneapolis, Minnesota. Retrieved July 11, 2009, from http://www.bowdoin.edu/~sputnam/rothbart-temperament-questionnaires/pdf/lesa-ellis-srcd-poster-reprint.pdf

Enfedaque, J. (1990). De los números a las letras [From numbers to letters]. *Suma, 5,* 23-34.

Engec, N. (2006). Relationship between mobility and student performance and behavior. *Journal of Educational Research, 99,* 167-178.

Engel, A. (1970). Teaching probability in intermediate grdes. In L. Råde (Ed.), *The teaching of probability and statistics* (pp. 87-150). Stockholm: Almqvist & Wiksell Förlag AB.

Esposito Lamy, C. (2003). Academic achievement in the context of poverty: Examining the effects of family risk on children's math and reading achievement over the elementary school years. *Dissertation Abstracts International-A, 64*(04), 1172. (UMI No. AAT 3088472)

Evans, J., & Tsatsaroni, A. (2000). Mathematics and its publics: Texts, contexts and users. *Social Epistemology, 14,* 55-68.

Falk, R. (1981). The perception of randomness. In C. Comiti, *Psychology of mathematics education. Proceedings of the fifth conference of the international group* (pp. 222-229). Seoul, South Korea. (ERIC Document Reproduction Service No. ED 225809)

Falk, R. (1992). A closer look at the probabilities of the notorious three prisoners. *Cognition, 43,* 197-223.

Falk, R., & Konold, C. (1994). Random means hard to digest. *Focus on Learning Problems in Mathematics, 16,* 2-12.

Falk, R., & Lann, A. (2008). The allure of equality: Uniformity in probabilistic and statistical judgment. *Cognitive Psychology, 57,* 293-334.

Fajemidagba, O. (1983). The relationships between Piagetian cognitive developmental stages of concrete and formal operations and achievement on mathematical ratio and proportion problems. (Doctoral dissertation, State University of New York at Albany). *Dissertation Abstracts International-A, 43*(11), 3497.

Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Education and Psychological Measurement, 53*, 357-382.

Fennema, E., & Sherman, J. (1977). Sex-related differences in mathematics achievement, spatial visualization, and affective factors. *American Educational Research Journal, 14*, 51-71.

Fennema, E., & Sherman, J. (1978). Sex related differences in mathematics achievement and related factors: A further study. *Journal for Research in Mathematics Education, 9*, 189-203.

Festinger, L. (1957). *Theorie der kognitiven Dissonanz [A theory of cognitive dissonance]*. Evanston, IL: Row, Peterson.

Fey, J. T. (1979). Mathematics teaching today: Perspectives from three national surveys. *Mathematics Teacher, 72,* 490-504.

Finn, C. E. (2005, January 26). Grading school standards: A tale of two studies. *National Review Online, 57*. Retrieved on June 29, 2009, from http://www.nationalreview.com/comment/finn200501260740.asp

Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? *Educational Studies in Mathematics, 15,* 1-24.

Fisher, K. (1988). The student-and-professor problem revisited. *Journal for Research in Mathematics Education, 19*, 260-262.

Ford, C. B. (1977). A study of the relationships among pupil perception of teacher social power base, teacher pupil control ideology and teacher concern level in teachers of secondary mathematics. (Doctoral dissertation, University of Houston, 1976). *Dissertation Abstracts International-A, 37*(08), 5051.

Fosnot, C. T., & Dolk, M. (2002). *Young mathematicians at work: Constructing fractions, decimals, and percents*. Portsmouth, NH: Heinemann.

Fraser, B. J. (1998). Classroom environment instruments: Development, validity, and applications. *Learning Environments Research, 1*, 7-33.

Freeman, M. (1997). *Math and science on a personal level*. Lexington, KY: University of Kentucky, Appalachian Rural Systemic Initiative.

Freudenthal, H. (1970). The aims of teaching probability. In L. Råde (Ed.), *The teaching of probability and statistics* (pp. 151-168). Stockholm: Almqvist & Wiksell Förlag AB.

Freudenthal, H. (1973). *Mathematics as an educational task*. Boston: Kluwer Academic.

Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures.* Boston: Kluwer Academic.

Fuson, K. C. (1998). Pedagogical, mathematical, and real-world conceptual-support nets: A model for building children's multidigit domain knowledge. *Mathematical Cognition, 4,* 147-186.

Fuson, K. C., Kalchman, M. S., & Bransford, J. D. (2005). Mathematical understanding: An introduction. In M. S. Donovan & J. D. Bransford (Eds.), *How students learn: History, mathematics, and science in the classroom* (pp. 217-256). Washington DC: National Academies Press.

Fuys, D. J., & Liebov, A. K. (1997). Concept learning in geometry. *Teaching Children Mathematics, 3,* 248-251.

Garcia-Coll, C. T., Kagan, J, & Reznick, J. S. (1984). Behavioral inhibition in young children. *Child Development, 55,* 1005-1019.

Gardner, H. (1987). The theory of multiple intelligences. *Annals of Dyslexia, 37,* 19-35.

Gardner, H. (1989). Beyond a modular view of mind. In W. Damon (Ed.), *Child development today and tomorrow* (pp. 222-239). San Francisco, CA US: Jossey-Bass.

Gardner, H. (2009). The five minds for the future. *School Administrator, 66,* 16-21.

Gardner, H., & Hatch, T. (1989). Multiple intelligences go to school: Educational implications of the theory of multiple intelligences. *Educational Researcher, 18,* 4-9.

Garner, R., & Alexander, P. A. (1989). Metacognition: Answere and unanswered questions. *Educational Psychologist, 24*, 143-158.

Garson, G. D. (2009). *Data imputation for missing values* [Website]. Raleigh, NC: North Carolina State University. Retrieved January 4, 2010, from http://faculty.chass.ncsu.edu/garson/PA765/missing.htm

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology*. (3rd ed.). Needham, MA: Allyn & Bacon.

Gleick, J. (1987). *Chaos*. New York: Viking.

Goffin, R. D. (1993). A comparison of two new indices for the assessment of fit of structural equation models. *Multivariate Behavioral Research, 28*, 205-214.

Goldman, J., & Gardner, H. (1989). Multiple paths to educational effectiveness. In D.K.Lipsky, A. Gartner, D. K. Lipsky, & A. Gartner (Eds.), *Beyond separate* *education: Quality education for all* (pp. 121-139). Baltimore, MD England: Paul H. Brookes Publishing.

Graham, A. T., & Thomas, M. O. J. (2000). Building a versatile understanding of algebraic variables with a graphic calculator. *Educational Studies in Mathematics, 41,* 265-282.

Gray, S. S., Loud, B. J., & Sokolowski, C. P. (2005, October). *Undergraduates' errors in using and interpreting variables: A comparative study.* Paper presented at the 27th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Roanoke, VA.

Green, D. R. (1982). *Probability concepts in 11-16 year old pupils.* Loughborough, England: Centre for Advancement of Mathematical Education in Technology, University of Technology. Retrieved on June 20, 2009, from http://www.data-archive.ac.uk/doc/1946/mrdoc/pdf/a1946uab.pdf

Green, D. R. (1983a). From thumbtacks to inference. *School Science and Mathematics, 83*, 541-551.

Green, D. R. (1983b). School pupils' probability concepts. *Teaching Statistics, 5,* 34-42.

Gregg, J. (1995). Discipline, control, and the school mathematics tradition. *Teaching and Teacher Education, 11,* 579-593.

Gresham, G., Sloan, T., & Vinson, B. (1997). *Reducing mathematics anxiety in fourth grade "at-risk" students*. Orlando, FL: University of Central Florida. (ERIC Document Reproduction Service No. ED417931)

Ham, S., & Walker, E. (1999). *Getting to the right algebra: The Equity 2000 initiative in Milwaukee Public Schools.* Retrieved January 30, 2009, from http://www.mdrc.org/publications/98/workpaper.html

Hatch, T. C., & Gardner, H. (1986). From testing intelligence to assessing competencies: A pluralistic view of intellect. *Roeper Review, 8,* 147-150.

Haver, W. E. (1978). Developing skills in college algebra – A mastery approach. *The Two-Year College Mathematics Journal, 9,* 282-287.

Hegedus, S. J., & Kaput, J. J. (2004, July). *An introduction to the profound potential of connected algebra activities: Issues of representation, engagement and pedagogy.* Paper presented at the annual meeting of the International Group for the Psychology of Mathematics Education, Bergin, Norway.

Henderson, R. W., & Landesman, E. M. (1995). Effects of thematically integrated mathematics instruction on students of mexican descent. *Journal of Educational Research, 88,* 290-300.

Hennequin, P. L. (1970). Teaching of probability and statistics in the French lycée. In L. Råde (Ed.), *The teaching of probability and statistics* (pp. 319-336). Stockholm: Almqvist & Wiksell Förlag AB.

Hiebert, J. (2003). What research says about the NCTM standards. In J.Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to "Principles and standards for school mathematics"* (pp. 5-23). Reston, VA: National Council of Teachers of Mathematics.

Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65-100). Reston, VA: National Council of Teachers of Mathematics.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K. C., Wearne, D., Murray, H. et al. (1997). *Making sense: Teaching and learning mathematics with understanding.* Portsmouth, NH: Heinemann.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371-404). Reston, VA: National Council of Teachers of Mathematics.

Hiebert, J., Stigler, J. W., Jacobs, J. K., Givvin, K. B., Garnier, H., Smith, M. et al. (2005). Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 video study. *Educational Evaluation and Policy Analysis, 27,* 111-132.

Hoffman, E. S., & Caniglia, J. (2009). In their own words: Good mathematics teachers in the era of NCLB. *Mathematics Teacher, 102,* 468-473.

Holliday, B., Cuevas, G. J., Moore-Harris, B., Carter, J. A., Marks, D., Casey, R. M. et al. (2005). *Algebra 2*. New York: Glencoe McGraw-Hill.

Howe, R. (2005). *Comments on NAEP algebra problems.* Washington, D. C.: Brookings Institution. Retrieved on May 15, 2009, from http://www.brookings.edu/brown/~/media/Files/Centers/bcep/AlgebraicReasoningConferenceHowe.pdf

Hu, L.-T., & Bentler, P. M. (1995). Evaluating model fit. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76-99). Thousand Oaks, CA: Sage.

Hughes, J., & Kwok, O. (2007). Influence of student-teacher and parent teacher relationships on lower achieving readers’ engagement and achievement in the primary grades. *Journal of Educational Psychology, 99*, 39-51.

Huinker, D., Doyle, L. H., & Pearson, G. E. (1995). *Landscape of mathematics and science education in Milwaukee. A study of the Milwaukee Public Schools*.

Milwaukee, WI: University of Wisconsin-Milwaukee, Center for Mathematics and Science Education Research. (ERIC Document Reproduction Service No. ED379160)

Hume, B. (1970). The introduction of probability and statistics on the pre-college level in Western Australia. In L. Råde (Ed.), *The teaching of probability and statistics* (pp. 337-340). Stockholm: Almqvist & Wiksell Förlag AB.

Ismail, N. A. (2009). Understanding the gap in mathematics achievement of Malaysian students. *Journal of Educational Research, 102,* 389-394.

Izsak, A. (2004). Students' coordination of knowledge when learning to model physical situations. *Cognition and Instruction, 22,* 81-128.

Jendraszek, P. A. (2008). Misconceptions of probability among future teachers of mathematics. *Dissertation Abstracts International-A*, *69*(01), 2008. (UMI No. AAT 3299270)

Jones, F. H. (1987). *Positive classroom discipline*. Columbus, OH: McGraw-Hill.

Jöreskog, K. G., & Sörbom, D. (1984). *LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods.* Mooresville, IN: Scientific Software.

Judson, T. W. (1999). Japan: A different model of mathematics education. *Contemporary Issues in Mathematics Education, 36,* 75-81.

K-12 Mathematics Curriculum Center. (2005). *Curriculum summaries.* Author.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3,* 430-454.

Kahneman, D., & Tversky, A. (1973a). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5,* 207-232.

Kahneman, D., & Tversky, A. (1973b). On the psychology of prediction. *Psychological Review, 80,* 237-251.

Kahneman, D., & Tversky, A. (1982). Subjective probability: A judgment of representativeness. In D. Kahneman, P. Slovic, & A. Tversky, *Judgement under uncertainty: Heuristics and Biases* (pp. 32-48). Cambridge: University Press.

Kahneman, D., & Tversky, A. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293-315.

Kalchman, M. S., & Koedinger, K. R. (2005). Teaching and learning functions. In M. S. Donovan & J. D. Bransford (Eds.), *How students learn: History, mathematics, and science in the classroom* (pp. 351-396). Washington D.C.: National Academies Press.

Kant, E. (1781/1901). *Kritik der reinen Vernunft [Critique of pure reason]*. (J. M. D. Meiklejohn, Trans.). New York: P. F. Collier and Son.

Kaput, J. J., & Clement, J. (1979). Letter to the editor. *Journal of Children's Mathematics Behavior, 2,* 208.

Keirsey, D. (1998). *Please understand me II: Temperament, character, intelligence* (Revised ed.). Del Mar, CA: Prometheus Nemesis.

Kentucky Department of Education. (2006). *Combined curriculum document.* Frankfort, KY: Kentucky Department of Education.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Kieran, C. (1980, August). *The interpretation of the equals sign symbol for an equivalent relation vs. an operator symbol.* Paper presented at the Fourth International Conference for the Psychology of Mathematics Education, Berkely, CA.

Kieran, C. (1989). The early learning of algebra: A structural perspective. In S. Wagner & C. Kieran (Eds.), *Research issues in the learning and teaching of algebra* (pp. 33-56). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kieran, C. (1992). The learning and teaching of school algebra. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 390-419). Reston, VA: National Council of Teachers of Mathematics.

Kieran, C. (2007). Learning and teaching of algebra at the middle school through college levels: Building meaning for symbols and their manipulation. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 707-762). Reston, VA: National Council of Teachers of Mathematics.

Kieran, C. (2008). *What do students struggle with when first introduced to algebra symbols?* Reston, VA: National Council of Teachers of Mathematics. Retrieved on June 22, 2009, from http://www.nctm.org/uploadedFiles/Research_News_and_Advocacy/Research/Clips_and_Briefs/Brief%20-%20What%20Can%20We%20Learn.pdf

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington D.C.: National Academy Press.

Konold, C. (1988). *Understanding Students' Beliefs about Probability.* Massachusetts University: Amherst Scientific Reasoning Research Institute.

Kozol, J. (1992). *Savage inequalities*. New York: Harper Perennial.

Kozol, J. (2005). *The shame of the nation*. New York: Crown Publishers.

Küchemann, D. (1978). Children's understanding of numerical variables. *Mathematics in School, 9*, 23-26.

Küchemann, D. (1981). Cognitive demand of secondary school mathematics items.

Educational Studies in Mathematics, 12, 301-316.

Lai-Yin, L., & Rob, R. (2005). Chinese and English probabalistic thinking and risk taking in gambling. *Journal of Cross-Cultural Psychology, 36,* 621-627.

Lamon, S. J. (1999). *Teaching fractions and ratios for understanding: Essential content knowledge and instructional strategies for teachers.* Philadelphia: Lawrence Erlbaum Associates.

Lamon, S. J. (2007). Rational numbers and proportional reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 629-667). Reston, VA: National Council of Teachers of Mathematics.

LaRocque, M. (2008). Assessing perceptions of the environment in elementary classrooms: The link with achievement. *Educational Psychology in Practice, 24,* 289-305.

Larson, R., Boswell, L., Kanold, T. D., & Stiff, L. (2010). *Algebra 2*. Austin, TX: Holt McDougal.

Larson, R., & Farber, B. (2009). *Elementary statistics: Picturing the world.* Upper Saddle River, NJ: Pearson Prentice Hall.

Le, C. (2005, January 6). Latest report praises school standards – Accountability, climate rate high. *The News Journal.* Retrieved June 29, 2009, from http://infoweb.newsbank.com/iw-search/we/InfoWeb?p_product=AWNB&p_theme=aggregated5&p_action=doc&p_docid=1077FD9EDD2F927B&d_place=NWJB&f_subsection=sLOCAL&f_issue=2005-01-06&f_publisher=

Lehrer, R., Strom, D., & Confrey, J. (2002). Grounding metaphors and inscriptional resonance: Children's emerging understanding of mathematical similarity.

Cognition & Instruction, 20, 359-398.

Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research, 60,* 1-64.

Leitzel, J. R. (1989). Critical considerations for the future of algebra instruction or a reaction to: "Algebra: what should we teach and how should we teach it?". In S.Wagner & C. Kieran (Eds.), *Research issues in the learning and teaching of algebra* (pp. 25-32). Hillsdale, NJ: Lawrence Erlbaum Associates.

Leitzel, J. R., Ellis, W., Enneking, M., Heid, M. K., & Jones, P. L. (1991). *A call for change: Recommendations for the mathematical preparation of teachers of mathematics. An MAA report.* Washington, D.C.: Mathematical Association of America. (ERIC Document Reproduction No. ED412109)

Leiva, M. A., Coes, L., & Brown, R. G. (2006). *Algebra 1: Explorations and applications*. Boston: Houghton Mifflin.

Lin, X., Schwartz, D. L., & Hatano, G. (2005). Toward teachers' adaptive metacognition. *Educational Psychologist, 40,* 245-255.

Liu, Y., & Thompson, P. (2007). Teachers' understandings of probability. *Cognition & Instruction, 25,* 113-160.

Lopez, E. M., Gallimore, R., Garnier, H., & Reese, L. (2007). Preschool antecedents of mathematics achievement of Latinos: The influence of family resources, early literacy experiences, and preschool attendance. *Hispanic Journal of Behavioral Sciences, 29,* 456-471.

Ma, X., Ma, L., & Bradley, K. D. (2008). Using multilevel modeling to investigate school effects. In A. A. O’Connell & D. B. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 59-110). Charlotte, NC: Information Age Publishing.

MacGregor, M., & Stacey, K. (1993). Cognitive models underlying students' formation of simple linear equations. *Journal for Research in Mathematics Education, 24*, 217-232.

MacGregor, M., & Stacey, K. (1997). Students' understanding of algebraic notation: 11-15. *Educational Studies in Mathematics, 33*, 1-19.

Maher, F. A., & Tetrault, M. K. T. (2001). *The feminist classroom* (2nd ed.). Oxford: Rowman & Littlefield.

Manouchehri, A., & Goodman, T. (1998). Mathematics curriculum reform and teachers: Understanding the connections. *Journal of Educational Research, 92*, 27-41.

Mathematics, Science, & Technology Education. (2005). *The Cliff-Hanger.* Champagne, IL: University of Illinois. Retrieved July 1, 2009, from http://mste.illinois.edu/activity/cliff/

Mathews, S. M. (1997). The effect of using two variables when there are two unknowns in solving algebraic word problems. *Mathematics Education Research Journal, 9,* 122-135.

McAvoy, M. R. (2001). Essays in country banking: Case studies of the First National Bank of Bloomington (Illinois). *Dissertation Abstracts International-A, 62*(06), 2198. (UMI No. AAT 3017161)

McDermott, L. C., Rosenquist, M. L., & Van Zee, E. H. (1987). Student difficulties in connecting graphs and physics: Examples from kinematics. *American Journal of Physics, 55,* 503-513.

McGatha, M. B., Bush, W. S., & Rakes, C. R. (2009). The effects of professional development in formative assessment on mathematics teaching performance and student achievement. *Journal of Multidisciplinary Evaluation*, 6, 32-43.

McKinley, J. E. (1960). Relationship between selected factors and achievment in a unit on probability and statistics for twelfth grade students. (Doctoral dissertation, University of Pittsburgh, 1960). *Dissertation Abstracts*, *21*(03), 561.

McLeod, D. (1992). Research on affect in mathematics education: A reconceptualization. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 575-596). Reston, VA: National Council of Teachers of Mathematics.

McMahon, A. J. (1979). The relationship between the pupil control ideology of mathematics teachers and student attitude toward mathematics. *Dissertation Abstracts International-A*, *40*(05), 2524. (UMI No. AAT 7923654)

McMahon, S. D., Wernsman, J., & Rose, D. S. (2009). The relation of classroom environment and school belonging to academic self-efficacy among urban fourth- and fifth-grade students. *Elementary School Journal, 109,* 267-281.

Mendick, H. (2008). Subtracting difference: Troubling transitions from GCSE to AS-level mathematics. *British Educational Research Journal, 34,* 711-732.

Ministry of Education (2007). *Secondary mathematics syllabuses*. Singapore: Author. Retrieved on March 28, 2009 from http://www.moe.gov.sg/education/syllabuses/sciences/files/maths-secondary.pdf

Mitchell, C. E. (1990). Seven areas neglected in school math. *Education Digest, 56,* 64-67.

Mlodinow, L. (2008). *The drunkard's walk: How randomness rules our lives*. New York: Pantheon.

Monaghan, F. (2000). What difference does it make? Children's views of the differences between some quadrilaterals. *Educational Studies in Mathematics, 42,* 179-196.

Morgan, E., & Huebner, A. (2008). *Adolescent growth and development* (Publication No. 350-850). Blacksburg: Virginia Polytechnic Institute and State University. Retrieved on March 26, 2009, from http://www.ext.vt.edu/pubs/family/350-850/350-850.pdf

Moschkovich, J. N. (1998). Resources for refining mathematical conceptions: Case studies in learning about linear functions. *Journal of the Learning Sciences, 7,* 209-237.

Moss, J. (2005). Pipes, tubes, and beakers: New approaches to teaching the rational-number system. In M. S. Donovan & J. D. Bransford (Eds.), *How students learn: History, mathematics, and science in the classroom* (pp. 309-350). Washington D.C.: National Academies Press.

Moss, J., Beatty, R., Barkin, S., & Shillolo, G. (2008). "What is your theory? What is your rule?" Fourth graders build an understandingof functions through patterns and generalizing problems. In C. E. Greenes & R. Rubenstein (Eds.), *Algebra and algebraic thinking in school mathematics: Seventieth yearbook.* (pp. 155-168). Reston, VA: National Council of Teachers of Mathematics.

Mueser, P. R., & Granberg, D. (1999). *The Monty Hall Dilemma revisited: Understanding the interaction of problem definition and decision making.* Unpublished manuscript, University of Missouri – Columbia. Retrieved on June 18, 2009, from http://129.3.20.41/eps/exp/papers/9906/9906001.html

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin, 105,* 430-445.

Muraki, E., & Bock, D. (2002). *PARSCALE 4.1* [Computer Program]. Chicago: Scientific

Software International.

Murayama, K., & Elliot, A. J. (2009). The joint influence of personal achievement goals and classroom goal structures on achievement-relevant outcomes. *Journal of Educational Psychology, 101,* 432-447.

Myers, I. B. (1962). *The Myers-Briggs type indicator.* Palo Alto, CA: Consulting Psychologists Press.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics (2006). *Curriculum focal points for prekindergarten through Grade 8 mathematics: A quest for coherence.* Reston, VA: Author. Retrieved on September 5, 2008, from http://www.nctm.org/standards/content.aspx?id=270

National Sciences Foundation, Mathematical Sciences Education Board. (1995). *Mathematical preparation of the technical work force: Report of a workshop.* Washington, D. C.: National Academy Press. Retrieved on May 31, 2009, from http://www.nap.edu/catalog.php?record_id=9066#toc

Nelson, R. D., Joseph, G. G., & Williams, J. (1993). *Multicultural mathematics: Teaching mathematics from a global perspective*. Oxford: Oxford University Press.

Nemirovsky, R., & Ferrara, F. (2009). Mathematical imagination and embodied cognition. *Educational Studies in Mathematics, 70,* 159-174.

Nichols, J. D., & Miller, R. B. (1994). Cooperative learning and student motivation.

Contemporary Educational Psychology, 19, 167-178.

No Child Left Behind Act of 2002, 20 U.S.C. § 6301 et seq. (West, 2002). Retrieved on February 1, 2009, from http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf

Oberdorf, C. D., & Taylor-Cox, J. (1999). Shape up! *Teaching Children Mathematics, 5,* 340-345.

Osterman, K. F. (2000). Students' need for belonging in the school community. *Review of Educational Research, 70*, 323-367.

Palarea Medina, M. M. (1999). La adquisición del lenguaje algebraico: Reflexiones de una investigación [The aquisition of algebraic language: Reflections of an investigation]. *Números, 40,* 3-28.

Papps, K. L. (2008). Productivity spillovers within families and firms. (Doctoral dissertation, Cornell University, 2008). *Dissertation Abstracts International, 69*(02), 2008.

Parham, J. W. (1993). An analysis of the effects of tutoring on seventh-grade students engaged in the mastery of pre-algebra concepts. (Doctoral dissertation, University of Maryland College Park, 1993). *Dissertation Abstracts International-A*, *54*(11), 4021.

Pascal, B., & Fermat, P. (1654/2006). Fermat and Pascal on probability (V. Sanford, Trans.). In P. Tannery and C. Henry, *Oeuvres de Fermat* (pp. 288-314). New York: Simon & Schuster.

Patton, M. W. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Philipp, R. A. (1992). A study of algebraic variables: Beyond the student-professor problem. *Journal of Mathematical Behavior, 11*, 161-176.

Pólya, G. (1957). *How to solve it: A new aspect of mathematical method.* Princeton, NJ: Princeton University Press.

Posner, M. I., & Rothbart, M. K. (2007). *Educating the human brain.* Washington D.C.: American Psychological Association.

Presmeg, N. C. (1997, October). *A semiotic framework for linking cultural practice and classroom mathematics.* Paper presented at the meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Bloomington, IL.

Protection of Human Subjects, 45 C.F.R. pt. 46. (2002). Retrieved June 1, 2009, from http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm

Quinn, R. J. (2004). Investigating probabilistic intuitions. *Teaching Statistics: An International Journal for Teachers, 26,* 86-88.

Råde, L. (1970). Teaching of probability and statistics on the pre-college level in Sweden. In L. Råde (Ed.), *The teaching of probability and statistics* (pp. 341-346). Stockholm: Almqvist & Wiksell Förlag AB.

Rakes, C. R., Valentine, J. C., & McGatha, M. B. (2010). *Methods of instructional improvement in algebra: A systematic review and meta-analysis*. Manuscript under review.

Raudenbush, S. W. (2009). *Optimal Design* [Software]. Lincolnwood, IL: Scientific Software International.

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

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2009). *Hierarchical linear modeling 6.08 for Windows* [Computer Software]. Lincolnwood, IL: Scientific Software

International.

Resnick, L. B. (1983). Mathematics and science learning: A new conception. *Science, 220*, 477-478.

Resnick, M. D. (1987). *Choices: An introduction to decision theory.* Minneapolis: University of Minnesota Press.

Reys, B. (2006). *The intended mathematics curriculum as represented in state-level curriculum standards: Consensus or confusion?* Charlotte, NC: IAP.

Reys, B., & Lappan, G. (2007). Consensus or confusion? *Phi Delta Kappan, 88,* 676-680.

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., 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.

Rodgers, C. E. (1995). An investigation of two instructional methods for teaching selected pre-algebra concepts to minority at-risk seventh-grade mathematics students. *Dissertation Abstracts International-A*, *56*(08), 3042. (UMI No. AAT 9539938)

Ronau, R. N., Rakes, C. R., Wagener, L., & Dougherty, B. (2009, February). *A comprehensive framework for teacher knowledge: Reaching the goals of mathematics teacher preparation.* Paper presented at the annual meeting of the Association of Mathematics Teacher Educators, Orlando, FL.

Ronau, R. N., & Rakes, C. R. (2010). *A comprehensive framework for teacher knowledge: A lens for examining research and teaching.* Manuscript in preparation.

Ronau, R. N., Wagener, L., & Rakes, C. R. (2009, April). A comprehensive framework for teacher knowledge: A lens for examining research. In R. N. Ronau (Chair), *Knowledge for Teaching Mathematics, a Structured Inquiry.* Symposium conducted at the annual meeting of the American Educational Research Association, San Diego, CA.

Rosnick, P., & Clement, J. (1980). Learning without understanding: The effect of tutoring strategies on algebra misconceptions. *Journal of Mathematical Behavior, 3*, 3-27.

Rothbart, M. K. (1994a). Broad dimensions of temperament and personality. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 337-341). New York: Oxford University Press.

Rothbart, M. K. (1994b). Emotional development: Changes in reactivity and self-regulation. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 369-375). New York: Oxford University Press.

Rothbart, M. K. (2006). *The Early Adolescent Temperament Questionnaire – Revised (EATQ-R)*. Retrieved July 11, 2009, from Bowdoin College, Mary Rothbart Temperament Lab Web site: http://www.bowdoin.edu/~sputnam/rothbart-temperament-questionnaires/instrument-descriptions/early-adolescent-temperament.html

Rothbart, M. K., Ahadi, S. A., & Evans, D. E. (2000). Temperament and personality: Origins and outcomes. *Journal of Personality and Social Psychology, 78*, 122-135.

Rothbart, M. K., Ahadi, S. A., & Hershey, K. L. (1994). Temperament and social behavior in childhood. *Merrill-Palmer Quarterly, 40*, 21-39.

Rothbart, M. K., & Jones, L. B. (1998). Temperament, self-regulation, and education. *School Psychology Review, 27*, 479-492.

Rouche, N. (2003). *Un fil conducteur pour l'enseignement des mathematiques [A thread for teaching mathematics].* Nevilles, Belgium: Centre de Recherche sur l'Enseignement des Mathématiques [Center for Research on Teaching Mathematics]. Retrieved on July 9, 2009, from http://www.enseignement.be/download.php?do_id=2279&do_check=

Rousseau, D. M. (1976). *The relationship of task characteristics of attitudes, absenteeism, stress, and performance among college students.* Washington, DC: U. S. Department of Health, Education, & Welfare National Institute of Education. (ERIC Document Reproduction Service No. ED140181)

Rousseau, J.-J. (1979). *Emile: or, On education.* (A. Bloom, Trans.). New York: Basic Books. (Original work published 1762)

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Hoboken, NJ: John Wiley & Sons.

Rudasill, K. M. (2009). *Temperamental attention and activity, classroom emotional support, and academic achievement in third grade.* Manuscript submitted for publication.

Rueda, M. R., Rothbart, M. K., Saccomanno, L., & Posner, M. I. (2007). Modifying brain networks underlying self-regulation. In D. Romer & E. F. Walker, *Adolescent psychopathology and the developing brain* (pp. 401-419). New York: Oxford University Press.

Saul, M. (2008). Algebra: The mathematics and the pedagogy. In C. E. Greenes & R. Rubenstein (Eds.), *Algebra and algebraic thinking in school mathematics* (pp.

63). Reston, VA: National Council of Teachers of Mathematics.

Schield, M. (2002). Statistical literacy inventory: Reading and interpreting tables and graphs involving rates and percentages. Minneapolis, MN: Augsburg College, W. M. Keck Statistical Literacy Project. Retrieved on June 19, 2009, from http://web.augsburg.edu/~schield/MiloPapers/StatLitKnowledge2r.pdf

Schield, M. (2005). *Five percentage table survey* (Version 2b). Retrieved July 9, 2009, from Augsburg College, W. M. Keck Statistical Literacy Project Web site: http://www.statlit.org/pdf/2005Schield5PctgTableSurvey.pdf

Schield, M. (2006). Statistical literacy survey results: Reading graphs and tables of rates and percentages. Minneapolis, MN: Augsburg College, W. M. Keck Statistical Literacy Project. Retrieved on June 19, 2009, from http://www.augsburg.edu/statlit/pdf/2006SchieldIASSIST.pdf

Schmidt, W., Houang, R., & Cogan, L. (2002). A coherent curriculum: The case of mathematics. *American Educator, 26*, 1-18.

Schmidt, W. H., McKnight, C. C., Raizen, S. A., Jakwerth, P. M., Valverde, G. A., Wolfe, R. G., Britton, E. D., Bianchi, L. J., & Houang, R. T. (1997). A splintered vision: An investigation of U.S. science and mathematics education. Executive summary. *Wisconsin Teacher of Mathematics, 48,* 4-9.

Schoenfeld, A. H. (1982, March). *Beyond the purely cognitive: Metacognition and social cognition as driving forces in intellectual performance.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Schoenfeld, A. H. (1985, April). *Students' beliefs about mathematics and their effects on mathematical performance: A questionnaire analysis.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Schoenfeld, A. H. (1989). Explorations of students' mathematical beliefs and behavior. *Journal for Research in Mathematics Education*, 20, 338-355.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334-371). Reston, VA: National Council of Teachers of Mathematics.

Schoenfeld, A. H., Smith, J. L., & Arcavi, A. (1993). Learning: The microgenetic analysis of one student's evolving understanding of a complex subject matter domain. In R. Glaser (Ed.), *Advances in Instructional Psychology* (Vol. 4, pp. 55-175). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*, 460-475.

Schreiber, J. B. (2000, October). *Scoring above the international average: A logistic regression model of the TIMSS advanced mathematics exam.* Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL.

Schumacker, R. E., Young, J. I., & Bembry, K. L. (1995). Math attitudes and achievement of Algebra I students: a comparative study of computer-assisted and traditional lecture methods of instruction. *Computers in the Schools, 11,* 27-33.

Secada, W. G. (1992). Race, ethnicity, social class, language, and achievement in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 623-660). Reston, VA: National Council of Teachers of Mathematics.

Selvin, S. (1975a). Letters to the editor: A problem in probability. *American Statistician,* *29*, 67-71.

Selvin, S. (1975b). Letters to the editor: On the Monty Hall problem. *American Statistician, 29*, 131-134.

Serra, M. (2003). *Discovering geometry: An investigative approach* (3rd ed.). Emeryville, CA: Key Curriculum Press.

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics, 22*, 1-36.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental design*. Boston; New York: Houghton Mifflin Company.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465-494). Reston, VA: National Council of Teachers of Mathematics.

Shaughnessy, J. M. (2003). Research on students' understandings of probability. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 216-226). Reston, VA: National Council of Teachers of Mathematics.

Shaughnessey, J. M. (2006). Research on students' understanding of some big concepts in statistics. In A. Flores, C. Pateracki, F. Bullard, J. M. Landwehr, & A. Hess (Eds.), *Thinking and reasoning with data and chance: 68th NCTM yearbook* (pp. 77-98). Reston, VA: National Council of Teachers of Mathematics.

Shaughnessy, J. M. (2007). Research on Statistics Learning and Reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957-1010). Reston, VA: National Council of Teachers of Mathematics.

Shaughnessey, J. M., & Bergman, B. (1993). Thinking about uncertainty: Probability and statistics. In P. S. Wilson (Ed.), *Research ideas for the classroom* (pp. 177-197). Reston, VA: National Council of Teachers of Mathematics.

Shaughnessy, J. M., & Burger, W. F. (1985). Spadework prior to deduction in geometry. *Mathematics Teacher, 78*, 419-428.

Shaughnessy, J. M., & Dick, T. (1991). Monty's dilemma: Should you stick or switch? *The Mathematics Teacher, 84*, 252-256.

Shermer, M. (2008). A random walk through middle land. *Scientific American, 299,* 40-41.

Shevlin, M., & Miles, J. N. V. (1998). Effects of sample size, model specification, and factor loadings on the GFI in confirmatory factor analysis. *Personality & Individual Differences, 25*, 85-90.

Shrewsbury, C. M. (1993). What is feminist pedagogy? *Women's Studies Quarterly, 3-4*, 8-15.

Shulte, A. P. (1967). The effects of a unit in probability and statistics on students and teachers of ninth-grade general mathematics. (Doctoral dissertation, University of Michigan, 1967). *Dissertation Abstracts - A*, *28* (12), 4962.

Sierpinska, A., & Kilpatrick, J. (1998). *Mathematics education as a research domain.* Kluwer Academic Publishers: Dordrecht.

Sinclair, R., & Ghory, W. (1979, April). *Curriculum as environments for learning: A practical meaning and model.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Silver, H. F., Brunsting, J. R., & Walsh, T. (2008). *Math tools grades 3-12: 64 ways to* *differentiate instruction and increase student engagement*. Thousand Oaks, CA: Corwin Press.

Silver, H., Strong, R., & Perini, M. (1997). Integrating learning styles and multiple intelligences. *Educational Leadership, 55,* 22-27.

Skemp, R. R. (1976/2006). Relational understanding and instrumental understanding. *Mathematics Teaching, 77,* 20-26. Reprinted in *Mathematics Teaching in the Middle School, 12*, 88-95.

Slavin, R. E., & Karweit, N. L. (1982). *Student teams and mastery learning: A factorial experiment in urban math nine classes.* Baltimore, MD: Johns Hopkins University, Center for Social Organization of Schools. (ERIC Document Reproduction Service No. ED215904)

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research, 78,* 427-515.

Slavin, R. E., Lake, C., & Groff, C. (2009). Effective programs in middle and high school mathematics: A best-evidence synthesis. *Review of Educational Research*, 79, 839-911.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences, 3,* 115-163.

Smith, R. J. (2003). Equal arcs, triangles, and probability. *Mathematics Teacher, 96,* 618-621.

Socas Robayna, M. M. (1997). Dificultades, obstáculos y errores en el aprendizaje de las matemáticas en la educación secundaria. [Difficulties, obstacles, and errors in learning mathematics in secondary education.]. In L. R. Romero (Ed.), *La* *educación matemática en la enseñanza secundaria. [Mathematics education at the secondary level.]* (pp. 125-154). Barcelona: Horsori y Instituto de Ciencias de la Educación.

Soper, D. (2004). *Statistics calculators*. Retrieved February 20, 2009, from http://www.danielsoper.com/statcalc/calc11.aspx.

Sowder, J. (1991). *Understanding as a Basis for Teaching: Mathematics and Science for Prospective Middle School Teachers. Final Report.* San Diego State University: Center for Research in Mathematics, Science, & Education. (ERIC Document Reproduction No. ED339590)

Spillane, J. P. (2000). Cognition and policy implementation: District policymakers and the reform of mathematics education. *Cognition and Instruction, 18,* 141-179.

Spybrook, J. (2008). Power, sample size, and design. In A. A. O'Connell & D. B. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 273-311). Charlotte, NC: Information Age Publishing.

Stacey, K. (1989). Finding and using patterns in linear generalising problems. *Educational Studies in Mathematics, 20,* 147-164.

Steen, L. A. (2007). How mathematics counts. *Educational Leadership, 65,* 8-14.

Steiger, J. H. (1998). A note on multiple sample extensions of the RMSEA fit index. *Structural Equation Modeling: A Multidisciplinary Journal, 5,* 411-419.

Steiger, J. H., & Lind, J. C. (1980, June). *Statistically-based tests for the number of common factors*. Paper presented at the Psychometric Society Annual Meeting, Iowa City, IA.

Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika, 50,* 253-264.

Stein, Z., Connell, M., & Gardner, H. (2008). Exercising quality control in interdisciplinary education: Toward an epistemologically responsible approach. *Journal of Philosophy of Education, 42,* 401-414.

Stein, M. K., Remillard, J., & Smith, M. S. (2007). How curriculum influences student learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 319-369). Reston, VA: National Council of Teachers of Mathematics.

Stemler, S. E., Elliott, J. G., Grigorenko, E. L., & Sternberg, R. J. (2006). There's more to teaching than instruction: Seven strategies for dealing with the practical side of teaching. *Educational Studies, 32,* 101-118.

Stevens, J. (2001). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Stibel, J. M., Dror, I. E., Ben-Zeev, T. (2009). The collapsing choice theory: Dissociating choice and judgment in decision making. *Theory and Decision, 66*, 149-179.

Stigler, J. W., & Hiebert, J. (1997). Understanding and improving classroom mathematics instructions: An overview of the TIMSS video study. *Phi Delta Kappan, 79,* 14-21.

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900.* Cambridge: Harvard University Press.

Stiff, L. V., Johnson, J. L., & Johnson, M. R. (1993). Cognitive issues in mathematics education. In P. S. Wilson (Ed.), *Research ideas for the classroom: High school mathematics* (pp. 3-20). Reston, VA: National Council of Teachers of Mathematics.

Stipek, D. (2006). Relationships matter. *Educational Leadership, 64*, 46-49.

Stone, J. R., Alfeld, C., & Pearson, D. (2008). Rigor and relevance: Enhancing high school students' math skills through career and technical education. *American Educational Research Journal, 45,* 767-795.

Stringfield, S. C., & Teddlie, C. (2004). The multiple roles of N.L. Gage in the creation of educational effectiveness fields. *Teaching and Teacher Education, 20,* 397-404.

Strong, R. W., Perini, M., Silver, H., & Thomas, E. (2004). Creating a differentiated mathematics classroom. *Educational Leadership, 61,* 73-78.

Strong, R. W., Silver, H. F., & Perini, M. J. (2001). Making students as important as standards. *Educational Leadership, 59,* 56-61.

Strutchens, M. (1995). *Multicultural mathematics: A more inclusive mathematics. ERIC Digest*. Washington D. C.: Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED380295)

Swaminathan, H. (2003). Bargaining and intra-household resource allocation: An analysis of the impact of credit and land in Malawi. *Dissertation Abstracts International-A, 64*(07), 2592. (UMI No. AAT 3098250)

Swanson, H. L. (1990). Influence of metacognitive knowledge and aptitude on problem solving. *Journal of Educational Psychology, 82*, 306-314.

Swenson, K. A. (1998). Middle school mathematics teachers' subject matter knowledge and pedagogical content knowledge of probability: Its relationship to probability instruction. *Dissertation Abstracts International-A, 59*(02), 440. (UMI No. AAT 9824767)

Swindal, D. N. (2000). Learning geometry and a new language. *Teaching Children Mathematics, 7,* 246-250.

Tankersley, M., Landrum, T. J., & Cook, B. G. (2004). How research informs practice in the field of emotional and behavioral disorders. In R. B.Rutherford, M. M. Quinn, & S. R. Mathur (Eds.), *Handbook of research in emotional and behavioral disorders* (pp. 98-116). New York: Guilford Press.

Tapia, M., & Marsh, G. E. (2004). An instrument to measure mathematics attitudes. *Academic Exchange Quarterly, 8*, 16-21.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.

Tenenbaum, G. (1986). The effect of quality of instruction on higher and lower mental processes and on the prediction of summative achievement. *Journal of Educational Research, 80,* 105-114.

Thomas, A., Chess, S., Birch, H. G., Hertzig, M. E., & Korn, S. (1963). *Behavioral individuality in early childhood.* New York: New York University Press.

Thorndike, E. L. (1913). *Educational psychology* (Vols. 1-2). New York: Columbia University Press.

Thorpe, J. A. (1989). Algebra: What should we teach and how should we teach it? In S.Wagner & C. Kieran (Eds.), *Research issues in the learning and teaching of algebra* (pp. 11-24). Hillsdale, NJ: Lawrence Erlbaum Associates.

Torigoe, E., & Gladding, G. (2006). Same to us, different to them: Numeric computation versus symbolic representation. *AIP Conference Proceedings, 883,* 153-156.

Tungate, M. (2008). Educators throughout state get 'thoughtful' about instruction. *Kentucky Teacher, August 2008,* 8-9.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105-110.

Urbina, S. (2004). *Essentials of psychological testing*. Hoboken, NJ: John Wiley & Sons.

U. S. Census Bureau. (2009). *Geography & environment: Environmental industry; Threatened and endangered species.* Washington, D. C.: The 2009 Statistical Abstract, U. S. Fish & Wildlife Service. Retrieved on May 14, 2009, from http://www.census.gov/compendia/statab/cats/geography_environment/environmental_industry_threatened_and_endangered_species.html

U. S. Department of Education. (1996). *National Assessment of Educational Progress (NAEP) questions*. Retrieved on July 10, 2009, from http://nces.ed.gov/nationsreportcard/itmrls/.

U. S. Department of Education. (2005). *National Assessment of Educational Progress (NAEP) questions*. Retrieved on November 3, 2008, from http://nces.ed.gov/nationsreportcard/itmrls/

U. S. Department of Education. (2007). *National Assessment of Educational Progress (NAEP) questions*. Retrieved on November 3, 2008, from http://nces.ed.gov/nationsreportcard/itmrls/

U. S. Department of Education. (2008a). *National Assessment of Educational Progress (NAEP) questions*. Retrieved on November 3, 2008, from http://nces.ed.gov/nationsreportcard/tdw/item_development/develop.asp

U. S. Department of Education. (2008b). *Overall descriptive statistics for item blocks, as defined after scaling, grade 4 mathematics combined national and state main assessment: 2003.* Retrieved on May 26, 2009, from http://nces.ed.gov/nationsreportcard/tdw/analysis/2002_2003/initial_classical_descrip2003mathg4.asp

U. S. Department of Education. (2008c). *Overall descriptive statistics for item blocks as* *defined after scaling, grade 8 mathematics combined national and state main assessment: 2003.* Retrieved on May 26, 2009, from http://nces.ed.gov/nationsreportcard/tdw/analysis/2002_2003/initial_classical_descrip2003mathg8.asp

U. S. Department of Education. (2008d). *Overall descriptive statistics for item blocks as defined after scaling, grade 12 mathematics national main assessment (accommodations not permitted [R2]): 2000.* Retrieved on May 26, 2009, from http://nces.ed.gov/nationsreportcard/tdw/analysis/2000_2001/initial_classical_descrip2000mathg12nonacc.asp

U. S. Department of Education. (2008e). *Overall descriptive statistics for item blocks as defined after scaling, grade 12 mathematics national main assessment (accommodations permitted [R3]): 2000.* Retrieved on May 26, 2009, from http://nces.ed.gov/nationsreportcard/tdw/analysis/2000_2001/initial_classical_descrip2000mathg12.asp

Usher, E. L. (2009). Sources of middle school students' self-efficacy in mathematics: A qualitative investigation. *American Educational Research Journal, 46,* 275-314.

Van De Walle, J. A. (2007). *Elementary and middle school mathematics*. (6th ed.). Boston: Pearson.

Van Dooren, W., De Bock, D., Depaepe, F., Janssens, D., & Verschaffel, L. (2003). The illusion of linearity: Expanding the evidence towards probabilistic reasoning. *Educational Studies in Mathematics, 53,* 113-138.

Van Hiele, P. M. (1959/1984a). A child's thought and geometry. In D. Fuys, D. Geddes, & R. Tischler (Eds.), *English translation of selected writings of Dina Van Hiele-Geldof and Pierre M. Van Hiele* (pp. 243-252). Washington, D. C.: National

Science Foundation. (ERIC Document Reproduction Service No. ED 287697)

Van Hiele, P. M. (1959/1984b). English summary by Pierre Marie Van Hiele of "The problem of insight in connection with school children's insight into the subject matter of geometry." In D. Fuys, D. Geddes, & R. Tischler (Eds.), *English translation of selected writings of Dina Van Hiele-Geldof and Pierre M. Van Hiele* (pp. 237-242). Washington, D. C.: National Science Foundation. (ERIC Document Reproduction Service No. ED 287697)

Van Langen, A., Rekers-Mombarg, L., & Dekkers, H. (2008). Mathematics and science choice following introduction of compulsory study profiles into dutch secondary education. *British Educational Research Journal, 34,* 733-745.

VanLehn, K. (1980). Bugs are not enough: Empirical study of bugs, impasses, and repairs in procedural skills. *Journal of Mathematical Behavior, 3,* 3-71.

VanLehn, K. (1983). On the representation of procedures in repair theory. In H. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 201-253). New York: Academic Press.

Vogel, C. (2008). Algebra: Changing the equation. *District Administration, 44,* 34-40.

Von Glasersfeld, E. (1987). Learning as a constructive activity. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 3-18). Hillsdale, NJ: Lawrence Erlbaum.

Von Minden, A. M., Walls, R. T., & Nardi, A. H. (1998). Charting the links between mathematics content and pedagogy concepts: Cartographies of cognition. *The Journal of Experimental Education, 66,* 339-358.

Vos Savant, M. (1990, September 9). "Ask Marilyn" column, *Parade Magazine*, p. 16.

Wagner, S., & Parker, S. (1993). Advancing algebra. In P. S. Wilson (Ed.), *Research* *ideas for the classroom* (pp. 119-140). Reston, VA: National Council of Teachers of Mathematics.

Watson, J. M., & Kelly, B. A. (2005). The winds are variable: Student intuitions about variation. *School Science & Mathematics, 105,* 252-269.

Watson, J. M., Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology, 34,* 1-29.

Watson, J. M., & Shaughnessy, J. M. (2004). Proportional reasoning: Lessons from research in data and chance. *Mathematics Teaching in the Middle School, 10,* 104-109.

Wei, F. Y. F., & Hendrix, K. G. (2009). Gender differences in preschool children's recall of competitive and noncompetitive computer mathematics games. *Learning, Media and Technology, 34,* 27-43.

Weinberg, A. (2007, October). *New perspectives on the student professor problem.* Paper presented at the 29th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (pp. 164-170), Reno, NV.

Welch, W. (1978). Science education in Urbanville. In R. Stake & J. Easley (Eds.), *Case studies in science education* (pp. 5-1-5-33). Urbana: University of Illinois.

Whicker, K. M., Bol, L., & Nunnery, J. A. (1997). Cooperative learning in the secondary mathematics classroom. *Journal of Educational Research, 91,* 42-48.

White, C. W. (1974). A study of the ability of seventh-grade and eighth-grade students to learn basic concepts of probability and the relationship between achievement in probability and selected factors. (Doctoral dissertation, University of Pittsburgh, 1974). *Dissertation Abstracts International - A*, *35* (04), 1969.

Whiteside, D. T. (Ed.). (1972). *The mathematical papers of Isaac Newton* (Vol. 5). London: Cambridge University Press.

Wilkins, C. W. (1993). *Effects of using a problem solving approach to Algebra I instruction*. Mississippi State, MS: Mississippi State University. (ERIC Document Reproduction Service No. ED372934)

Wolf, F. L. (1962). *Elements of probability and statistics*. New York: McGraw-Hill.

Wothke, W. (1993). Nonpositive definite matrices in structural equation modeling. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 256-293). Newbury Park, CA: Sage.

Wu, H. (2005, September). Key mathematical ideas in Grades 5-8. Paper presented at the annual meeting of the *National Council of Teachers of Mathematics*, Anaheim, CA.

Yates, D. S., Moore, D. S., & Starnes, D. S. (2006). *The practice of statistics*. (3rd ed.). New York: W.H. Freeman.

Zawojewski, J. S., & Shaughnessey, J. M. (2000). Mean and median: Are they really so easy? *Mathematics Teaching in the Middle School, 5,* 436-440.

Zohar, A., & Gershikov, A. (2008). Gender and performance in mathematical tasks: Does the context make a difference? *International Journal of Science and Mathematics Education, 6,* 677-693.

APPENDIX A

Algebra Lesson 1: Statistical Structure

Situation:

This lesson is the first of a probability unit designed for high school algebra classes. In this lesson, students will examine the structure of descriptive statistics and the normal distribution at an introductory level.

This lesson is designed for a 90 minute block period class.

Objectives

1. Students will describe the structure of data analysis.
2. Students will interpret mean, median, mode, variance, and standard deviation.
3. Students will construct normal distribution data displays.
4. Students will interpret the normal distribution data display.

Connections

In this lesson, probability is introduced as an extension of data analysis and the need to make inferences about a population. The opener begins this sequence by having students explore the notion of equality as it relates to rational numbers, a foundational concept to probability. In the 2nd lesson of this unit, the topics discussed in the present lesson will be reviewed further.

Materials

1) Measuring tapes to measure height in inches
2) LCD Projector
3) Microsoft PowerPoint (And Clicker, if available)
4) PowerPoint Presentation
5) Student Lesson Worksheet

KY Core Content 4.1 Standards

MA-HS-1.1.1: Students will compare real numbers using order relations (less than, greater than, equal to) and represent problems using real numbers.

MA-HS-4.1.1: Students will analyze and make inferences from a set of data with no more than two variables, and will analyze problems for the use and misuse of data representations.

MA-HS-4.1.2: Students will construct data displays for data with no more than two variables.

MA-HS-4.2.1: Students will describe and compare data distributions and make inferences from the data based on the shapes of graphs, measures of center (mean, median, mode) and measures of spread (range, standard deviation).

MA-HS-4.2.2: Students will know the characteristics of the Gaussian normal distribution (bell-shaped curve).

Procedures

1) Opener – Think, Pair, Share (@ 10 minutes overall)
    a. Draw a picture to explain why 1/5 is equal to 0.2
        i. Give 3-5 minutes to draw pictures. *(Teacher takes roll and posts)*
        ii. Pair up and discuss answers (@ 1-2 minutes)
        iii. Share out in pairs (2 minutes)
    b. Discuss sample answers on PowerPoint slide. (@ 2 minute)
        i. Two possible pictures are given.
2) Guided Notes (75 minutes) – Pass out Student Lesson Worksheet.
    a. Structure of Statistics: "The Statistical Pyramid" (Which is actually a ziggurat)
        i. The purpose of statistics is to say something about a population (Inference).
            1. Can't measure a population directly because of constraining factors (e.g., time, money, ability to identify all subjects of population)
            2. Instead, we have to estimate population "parameters" from samples.
        ii. Sample statistics are the foundation of all data analysis.
            1. Not particularly interesting by themselves (e.g., I can easily measure the attitudes of students in a classroom about a topic, but what I really want to measure is the attitude held by all teenagers in the U. S.)
            2. Descriptive Statistics: We either describe sample behavior by the center or its spread. The center statistics are mean, median, and mode. The spread statistics are variance, standard deviation, and range.
            3. The frequency with which we observe particular values in a sample is sometimes called either the "sample" distribution or "data" distribution. (We'll typically use "data" distribution to distinguish it from the "sampling" distribution)
        iii. Population parameters are also described by center and spread
            1. Center is typically the mean (although sometimes the median is used)
            2. Spread is variance and standard deviation.
            3. The frequency with which we observe particular values in a population is the "population distribution"
        iv. In between the population and data distributions is the pyramid staircase, the sampling distribution.
            1. What is a sampling distribution? We take a sample of N subjects, and compute a mean. The mean becomes a data point in the sampling distribution. We repeat this process a certain number of times, each time placing the sample mean into the sampling distribution set. When we have an "infinite" number of samples (of equal size), we have the

sampling distribution. (The frequency with which mean values are observed in repeated sampling)

2. The sampling distribution also has a center and a spread. The center is the mean (which is actually a mean of means) and standard deviation. The standard deviation of a sampling distribution is usually referred to as "standard error."

b. Symbols commonly used in data analysis:
   i. These symbols are organized by which distribution they belong to:
      1. In the sample,
         (a) x = observation data
         (b) $\bar{x}$ = sample mean
         (c) s^2 = sample variance
         (d) s = sample standard deviation
      2. In the sampling distribution,
         (a) $\mu_{\bar{x}}$ = sampling distribution mean (Mean of Means)
         (b) $\sigma_{\bar{x}}$ or SE = standard error
      3. In the population distribution,
         (a) μ = population mean
         (b) σ^2 = population variance
         (c) σ = population standard deviation
      4. Some general symbols used commonly:
         (a) Δ = Change (Looks like a triangle, but it's actually a Greek Capital Delta)
         (b) Σ = Sum (Greek Capital Sigma)
         (c) *df* = Degrees of Freedom

c. Describing data by the center
   i. Why would we want to do this? The center value can sometimes be a good representation of the values in the data set. Much easier to use one number instead of a thousand.
   ii. Graphing the data on a dotplot (a number line with each repeated point stacked)
   iii. Using the sample data set, go through computing the mean, median, and mode. Place a mark on the dotplot to show the mean, median, and mode.
   iv. Which center best represents this data set? In this case, the mean does a better job of representing the set (the 50 is high, but not high enough to be an outlier, and the median doesn't account for its high value).
   v. How to decide which center to use:
      1. Mean: Most commonly used center. Used when data are distributed "normally" (bell shaped curve).
      2. Median: Used when data set contains outliers.
      3. Mode: Used when all the data cluster around a single value.

d. Degrees of Freedom (df)

i. Defined as the number of independent observations in a population represented by a sample.

ii. What do we mean by independent observations?

1. Consider a sample of 4 people who are measured on some "score." The sample mean $\bar{x} = 20$, which means the sum of the scores was 80.
2. We automatically estimate the population mean μ to also be 20.
3. So, when we go to the next sample of 4 people from the population, the first three people's scores can be whatever they want (i.e., "free")
4. However, the fourth observation must make up the difference to get the sum to be 80 so that the population mean is still 20. We therefore say that it is "fixed" to track the population mean.
5. So, we say that every time we estimate a population parameter, we lose a degree of freedom. So, df = n – 1 (usually).

iii. Degrees of freedom are important. Most statistical calculations assume that one observation doesn't influence another (i.e., "independent")

iv. So, when we talk about samples, we think about sample size. But, when we talk about populations, we think about degrees of freedom.

e. Divide students into groups of 6 (or allow them to group themselves).

i. Measure heights in inches.

ii. Have students compute the mean and subtract the mean from each X.

iii. Have students add the second column. (If they don't get 0, then there are either rounding errors or computation errors)

iv. If we are interested in finding an average distance from the mean, why is this sum a problem? Because it means that regardless of the distances in the set, the average will always appear to be 0.

v. Why is the sum always 0? Difficult to tell from the table, so let's look at a sample set of data and a number line. Notice that the negative distances (observations below the mean) cancel out the positive distances (observations above the mean). Check your data set and see if it's true for yours as well!

vi. Computing the average distance from the mean can be thought of like calculating the distance between any other two points

1. Use Pythagorean Theorem
2. Notice that $a + b \neq c$ (and can't for any triangle)
3. But $a^2 + b^2$ does equal c^2. When we talk about a^2, b^2, or c^2, we are talking about area of squares.
4. Benefit of using squares: (1) Can be used to compute distance (point out that the Pythagorean Theorem re-

worked is the distance formula); (2) Eliminates the 0 sum problem.

5. How does it eliminate 0 sum problem? The square of any number is positive (positive • positive = positive; negative • negative = positive).

vii. Have students return to their height activity.
   1. Plot points on number line
   2. Draw in mean distances
   3. Draw the squares.
   4. Fill in the third column by squaring each mean distance.
   5. Find the sum of the third column and divide by 5 (Divide by 5 and not 6 because we want average distance in population, not sample, so use df instead of *n*).
   6. Go through sample data and allow students to follow their work with the sample.

viii. What does that get us?
   1. Average square area for the mean distances is the "variance"
   2. It's formula is: $Variance = \sigma^2 = \dfrac{\sum\left(X - \overline{X}\right)^2}{n-1}$
   3. It should be interpreted as the average amount of "noise" around the mean. It is the amount of data not represented by the mean.
   4. The side length of the variance square is the "standard deviation."
   5. It's formula is:

$$Standard\ \ Deviation = \sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum\left(X - \overline{X}\right)^2}{n-1}}$$

ix. Don't worry about memorizing formulas! Graphing calculator computes these values easily.
   1. Stat → Edit → Edit takes students to the lists to type in their data (have them use their data set while you go through steps)
   2. 2^{nd} → Quit returns them to the main calculator screen.
   3. Stat → Calc → 1 Variable Stats provides:
      (a) Mean
      (b) Sum of X
      (c) Sum of X^2
      (d) Population Standard Deviation
      (e) Sample Standard Deviation
      (f) Sample Size
      (g) Minimum X
      (h) Q1 (25^{th} percentile)
      (i) Median (50^{th} percentile)
      (j) Q3 (75^{th} percentile)

(k) Maximum X

x. Normal Distribution

1. The normal distribution is a frequency distribution that corresponds to the probability of seeing a range of values in the population. Simplest, most common way of getting to population inference from a sample.
2. Probability patterns are non-linear. (Rule of 3/Linear Proportions will not work to find unknowns).
3. Look at Normal Distribution equation. No need to memorize the formula. Just notice that this distribution is based on mean, variance, and standard deviation.

xi. 68-95-99 rule

1. 68% of population falls between $-\sigma$ and $+\sigma$. 95% of population falls between -2σ and $+2\sigma$. 99% of population falls between -3σ and 3σ WHEN DATA ARE NORMAL.
2. These percentages are approximations.
3. Finding critical values: add/subtract σ to/from mean.
4. Making inferences with normal data.

xii. Plotting normal curves

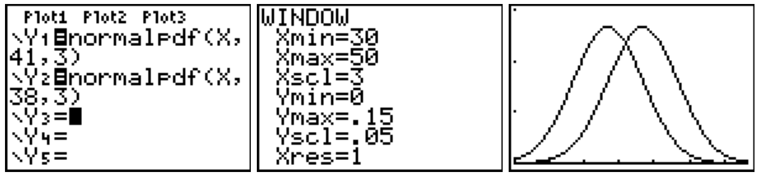
1. Graphing calculator command: Y = Normalpdf(X, mean, st. dev.)
2. Set Window.
   (a) XMin: A little lower than -3σ.
   (b) XMax: A little higher than 3σ.
   (c) XScl: Your choice. I tend to prefer σ.
   (d) YMin: 0
   (e) YMax: Your choice. I would think not higher than 0.3
   (f) YScl: Your choice. 0.1 is usually pretty good
   (g) XRes: Keep at 1.
3. Effect of standard deviation on curve height/width.
   (a) Higher SD = Shorter, wider graph
   (b) Lower SD = Taller, narrower graph
4. Effect of mean on curve. Shifts position.

3) Closure (5 minutes)

a. Give a couple of minutes for individual work.
b. Report out with whole class.
c. Sample Solution:

A news report posts that a political candidate has a 41% approval rating while her opponent has a 38% approval rating (SD = 3%). How does the normal distribution indicate that neither opponent is actually winning?

The majority of the normal distributions for each candidate overlap:



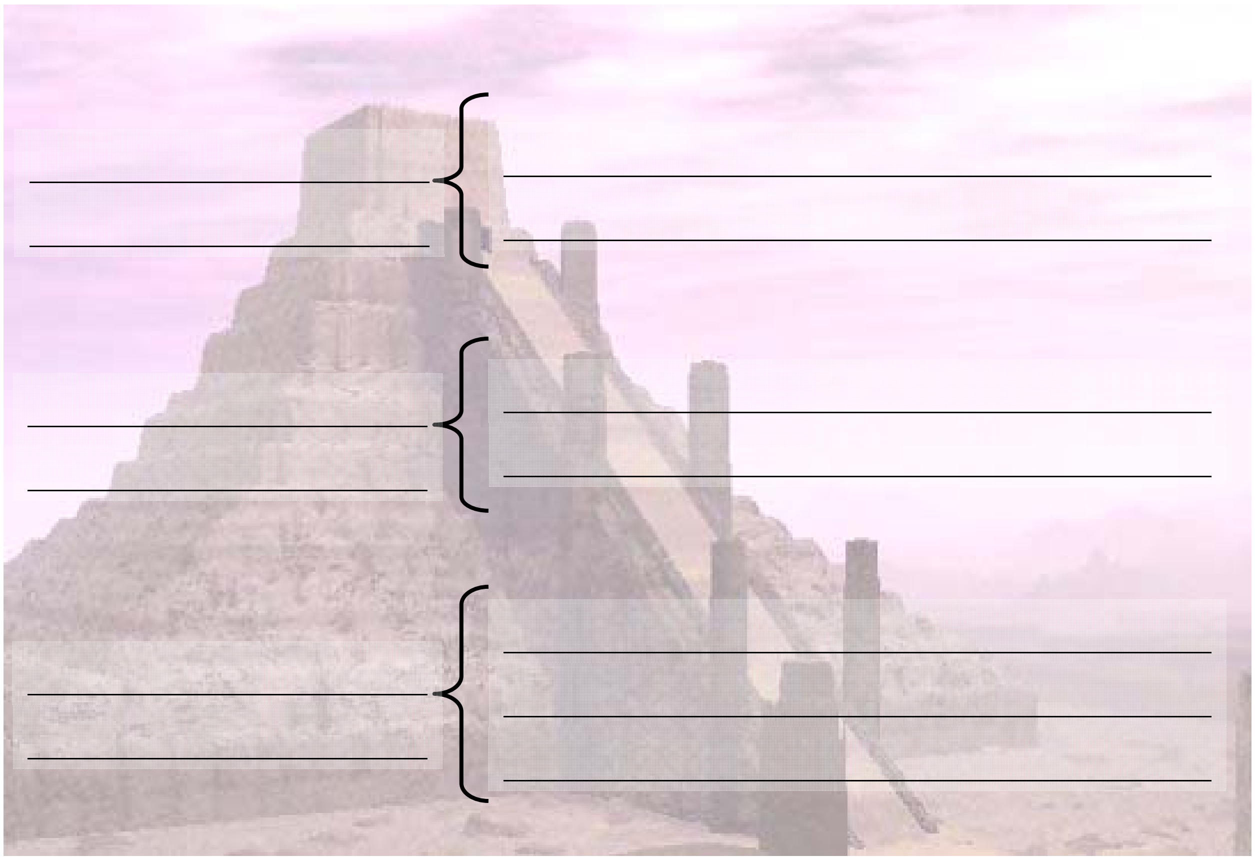
Assessment

1) Opener: Students connect rational number representations to notion of equality (1.1.1)
2) Guided Notes and In-Class Simulation (Teacher observes and questions while students work individually; Questioning during whole class discussion) (4.1.1, 4.1.2, 4.2.1, 4.2.2)
3) Students will do more practice in Lesson 2 with normal distribution (4.1.1, 4.1.2, 4.2.1, 4.2.2)

APPENDIX B

Algebra Lesson 1 Student Worksheet

1. The structure of statistics



2. Some symbols and their meaning:

Symbol	Meaning
Δ	
Σ	
df	
μ	
σ^2	
σ	
$\mu_{\bar{x}}$	
$\sigma_{\bar{x}}$	
x	
$\bar{x}$	
s^2	
s	

3. Describing data by the center
Sample Data Set: 50, 10,1, 7, 1, 25, 20
   a) Make a dotplot of the data.

   b) Compute the Mean. Mark it on the dotplot above.

   c) Compute the Median. Mark it on the dotplot above.

   d) Compute the Mode. Mark it on the dotplot above.

   e) Which center best represents this sample data set? Why?

4. What does the term "Degrees of Freedom" mean?

   a. How is it usually computed?

   b. Why is it important?

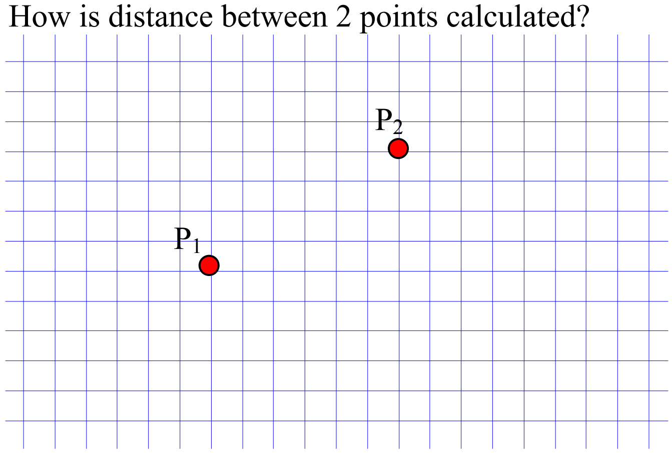
5. A) Collect the height of six people in the class (in inches).

X (Height)	X - $\bar{x}$	
1)		
2)		
3)		
4)		
5)		
6)		
$\bar{x}$ =	$\Sigma(X - \bar{x})$ =	

B) What happened when you tried to calculate the average distance to the mean?

C) Why do you think this happened?

6. How is distance between 2 points calculated?



7. What is the benefit of using square areas for distance?

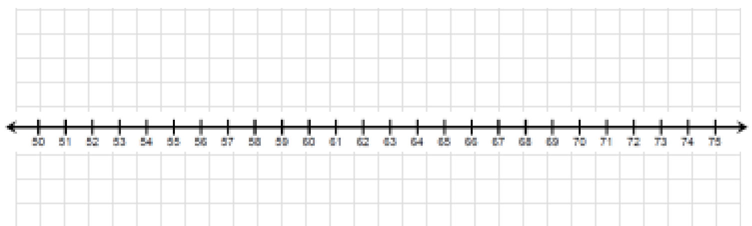
8. Plot your student height data and mean on the number line provided on the next page.

9. Compute the squares of the mean distances in the third column of the table in Problem 3; then draw the squares on the number line.

10. Find the sum of the squares from column 3 in problem 3.

    a. Divide the sum by 5. Why 5 and not 6?

    b. What side length will produce a square of that size?



11. The area of the average squared distance is referred to as…

    Formula:

    Meaning:

12. The side length of the average square is the…
    Another name for this length is…

    Formula:

    Meaning:

13. Graphing Calculator Operations

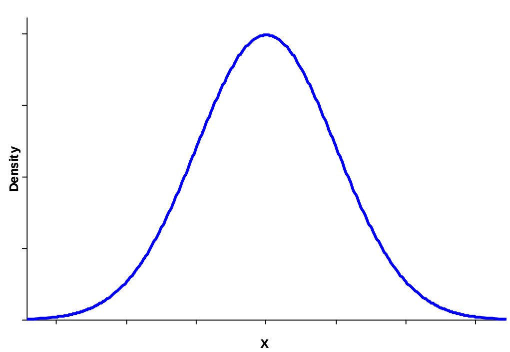
Description	Buttons	Purpose
STAT → ENTER	STAT ENTER	
2^{nd} → QUIT	QUIT 2nd MODE	
STAT → Calc → 1-Var Stats	STAT ENTER EDIT CALC TESTS	$\bar{x}$: $\sum x$: $\sum x^2$: S*x*: σ*x*: *n*: minX: Q1: Med: Q3: maxX:

14. What is the normal distribution? (*Read* *http://davidmlane.com/hyperstat/normal_distribution.html*)

15. Identify the mean, variance, and standard deviation in the mathematical formula:

$$\textbf{Height} \sim \mathbf{N}(\mu,\sigma) = \frac{\mathbf{e}^{\frac{-(\mathbf{x}-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

16. What is the 68 – 95 – 99 rule?



17. How do you compute the values for the X axis?

18. Complete the following for your example data.

   a. Approximately 68% of the population will fall between _____ and _____ inches.
   b. Approximately 95% of the population will fall between _____ and _____ inches.
   c. Approximately 99% of the population will fall between _____ and _____ inches.
   d. Approximately 84% of the population will be shorter than _____ inches.
   e. Approximately 0.5% of the population will be taller than _____ inches.
   f. Approximately 16% of the population will be shorter than _____ inches.
   g. Approximately 99.5% of the population will be taller than _____ inches.

19. Plotting Normal Distributions
   a. What is the command sequence for graphing a normal distribution on a graphing calculator?

   b. How do you determine the Window to set for a Normal Distribution?
      i. XMin:
      ii. XMax:
      iii. XScl:
      iv. YMin:
      v. YMax:
      vi. YScl:
      vii. XRes:

   c. Plot two normal distributions on the same graph. Graph A: ~N(25, 2); Graph B: ~N(25, 5).

   d. What is the effect of the mean on the graph of a normal distribution?

   e. What is the effect of a larger standard deviation on the normal distribution?

   f. What is the effect of a smaller standard deviation on the normal distribution?

APPENDIX C

Algebra Lesson 2: Randomness Lesson Plan

Situation:

This lesson is the second of a probability unit designed for high school algebra classes. In this lesson, students will examine randomness and sampling techniques used in experimentation. In subsequent lessons, students will examine counting principles and probability patterns of random variables.

This lesson is designed for a 90 minute block period class.

Objectives

1) Students will distinguish between patterns and randomness.
2) Students will predict ending positions of a random walk.
3) Students will distinguish between types of sampling patterns.

Connections

In the first lesson, students learned how to create normal distribution curves from means and standard deviations. In this lesson, students will explore patterns within random data and discover that the random data will follow a normal pattern. In the next lessons, students will build concepts of counting and probability on the foundation of randomness and normality.

Materials

6) Coins for flipping at Station 1: Determining Random Patterns and Station 2: Ant Walk
7) Computer with Internet for Station 3
    a. Go to Cliff Hanger applet or type in http://mste.illinois.edu/activity/cliff/.
    b. Note: The Cliff Hanger applet has sound; while sound is optional, it will make the station far livelier. ☺
8) LCD Projector
9) Microsoft PowerPoint (And Clicker, if available)
10) PowerPoint Presentation
11) Student Lesson Worksheet

KY Core Content 4.1 Standards

MA-HS-4.1.1: Students will analyze and make inferences from a set of data with no more than two variables, and will analyze problems for the use and misuse of data representations.

MA-HS-4.1.2: Students will construct data displays for data with no more than two variables.

MA-HS-4.2.1: Students will describe and compare data distributions and make inferences from the data based on the shapes of graphs, measures of center (mean, median, mode) and measures of spread (range, standard deviation).

MA-HS-4.2.2: Students will know the characteristics of the Gaussian normal distribution (bell-shaped curve).

MA-HS-4.3.1: Students will recognize potential for bias resulting from the misuse of sampling methods (e.g., non-random sampling, polling only a specific group of people, using limited or extremely small sample sizes) and explain why these samples can lead to inaccurate inferences.

MA-HS-4.3.2: Students will design simple experiments or investigations to collect data to answer questions of interest.

Procedures (88 minutes overall)

4) Opener – Think, Pair, Share (@ 8 minutes overall)
    a. Give 2-3 minutes to complete the work. *(Teacher takes roll and posts)*
    b. Pair up and discuss answers. (@ 2 minutes)
    c. Share out in pairs. (@ 2 minutes)
    d. Discuss sample answers on PowerPoint slide. (@ 2 minute)
        i. Review computation of critical values for normal distribution.
        ii. Review Probability Percentages and meaning of Percentile: 98th percentile does not mean that Ben got 98% of the questions correct; it means that he scored better than 98% of the other students who took the test.
5) **Pass out Student Lesson Worksheet.** Have students read about Randomness and Sampling in their textbook and answer as many questions on Page 1 as they can on their own. (5 minutes)
6) **Whole Class discussion of Page 1 Questions.** (15 minutes)
    a. What is randomness?
        i. Every outcome has an equal chance of being selected.
    b. Why is it important?
        i. Random patterns appear in the world in many places: atomic and molecular movement, lottery, decision under uncertainty
    c. Simple Random Sampling
        i. Every individual in a population has an equal and independent chance of being selected for the study. The sample is obtained through selection by chance, a table of random numbers, or computer-generated random numbers.
    d. Systematic Random Sampling
        i. Based on the number needed in the sample, every *n*th person in the target population is selected for the sample.
        ii. Used most often for product quality testing (e.g., every nth product)
        iii. If the order to be sampled is random, then no bias increase.
    e. Stratified Random Sampling

i. This (method) is used when the proportion of subgroups (strata) are known in the population; selection is random but from each of these strata.
ii. Used in political polling, voting (e.g., political districts form the strata)
iii. Especially useful when one or a few groups constitute a large portion of the population — in this type of case, the stratified sample can reduce bias, rather than increase it.

f. Convenience Sampling
i. Sampling is done on the basis of availability and ease of data collection rather than in terms of suitability based on research objectives/questions.
ii. Researchers (especially medical and social researchers) rely on willing participants (volunteers). This situation adds bias to the data — volunteers may share common traits that become over-represented in the sample.
iii. Randomly assigning volunteers to treatment/control group reduces this added bias.

7) Station Work (15 minutes at each station; 45 minutes total)
   a. Station 1: Distinguishing Randomness within sequences
   b. Station 2: Random Walk of Ants
   c. Station 3: Random Walk with Tourist at the Grand Canyon ("Cliff Hanger")
8) Return to normal seats. (@ 10 minutes)
   a. Compile frequency of outcomes for ant walk.
   b. Build a histogram of class data.
   c. If we did 100 trials, would the histogram change? Why or why not?
   d. If each trial consisted of 20 steps, would the histogram change? Explain. Note: PowerPoint Slide has histogram from two Monte Carlo samples.

9) Closure (5 minutes)
   What was the most surprising thing you learned today? Why did it surprise you?
   a. Write your own answer
   b. Discuss with a partner
   c. Share out with class

Assessment

4) Opener: Review of Normal Distribution (4.1.1, 4.2.1, 4.2.2)
5) Student Work Stations/Lesson Worksheet (4.1.2, 4.3.1, 4.3.2)

APPENDIX D

Algebra Lesson 2 Station Prompts

Station 1: Determining Random Patterns

1) A teacher asked Clare and Susan each to toss a coin a large number of times and to record every time whether the coin landed Heads or Tails. For each 'Heads,' a 1 is recorded and for each 'Tails,' a 0 is recorded. Here are the two sets of results:

CLARE	01011001100101011011010001110001101101010110010001
	01010011100110101100101100101100100101110110011011
	01010010110010101100010011010110011101110101100011
SUSAN	10011101111010011100100111001000111011111101010101
	11100000001000101001000000100011000101000000000011001
	00000001111100001101010010010011111101001100011000

a) Now one girl did it properly, by tossing the coin. The other girl cheated and just made it up. Which girl cheated? How can you tell?

b) Now try it yourself with a partner. One person flips the coin while the other records the outcome. Switch off every 10 flips. Flip the coin 100 times. Record the results below.

Flips	1	2	3	4	5	6	7	8	9	10
Flips 1-10										
Flips 11-20										
Flips 21-30										
Flips 31-40										
Flips 41-50										
Flips 51-60										
Flips 61-70										
Flips 71-80										
Flips 81-90										
Flips 91-100										

c) Does your simulation make you change your decision in Part A? Why or why not?

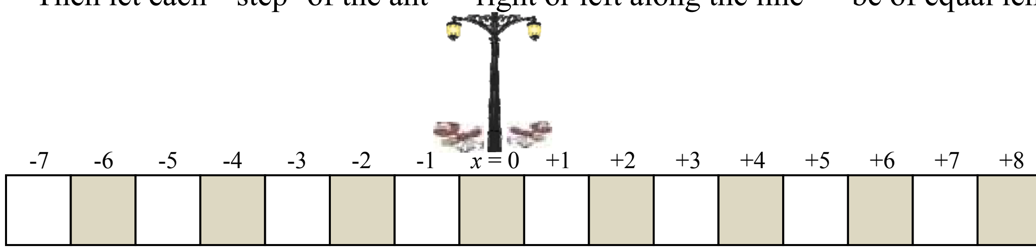
Station 2: Ants and Random Movement

Problem Statement:

If a wandering ant starts at a lamp post and takes steps of equal length along the street, how far will it be from the lamp post after a certain number, say N, steps? Though this question is seemingly trivial, it poses one of the most basic problems in statistical science.

It is easiest to visualize random motion (random walk) along one line, that is, in one dimension.

- Call x the position of the ant on a one-dimensional line.
- Locate the origin, that is x = 0, at the lamp post.
- Then let each ``step" of the ant — right or left along the line — be of equal length.



1) How far from the origin do you expect the wandering ant to end up after 10 steps?

2) After 10 steps is the ant more likely to be to the right or to the left of its starting point?

3) If another ant takes 10 steps from the starting point, then another ant, then another ant, what do you expect their average final position to be after 10 steps?

Simulation

- Choose the *direction* of the step the ant will take by flipping a coin:
  - (+1) If it is a head, the ant steps right and x increases by one.
  - (–1) If it is a tail, the ant steps left and x decreases by one.

4) What is the likelihood of getting a head or tail? What is the implication for the ant's steps?

5) Flip a penny ten times and move your ``ant" accordingly. Record the data in the table below. (Starting Position = 0, the lamp post)

Flip	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Result										
Ending Position										

6) Did the ending result surprise you? Why or why not?

7) Do the repeated trials for the ant walk represent random trials? Why or why not?

Whole Class work on Station 2 Ants

a. Compile the class data and construct a histogram.

Final Position of x	Frequency
-7	
-6	
-5	
-4	
-3	
-2	
-1	
0	
1	
2	
3	
4	
5	
6	
7	
8	

Now that you've had a chance to experiment, answer the following:

1) What would you say about the value of the average position of many random walkers?

2) If we conducted the same experiment with 100 trials of 10 steps, would the histogram be different? Why or why not?

3) If each trial had the ant walk 20 steps instead of 10, would the histogram be different from 100 trials of 10 steps? Explain.

Station 3: Cliff Hanger

Go to the Cliff Hanger applet at http://mste.illinois.edu/activity/cliff/.

Problem Statement:

A long day hiking through the Grand Canyon has discombobulated this tourist. Unsure of which way he is randomly stumbling, 1/3 of his steps are towards the edge of the cliff, while 2/3 of his steps are towards safety. From where he stands, one step forward will send him tumbling down. What is the probability that he can escape unharmed?

1) Build a factor tree for the possible ending points for the tourist's next 4 steps. (E = Toward Edge; S = Toward Safety). For each outcome, compute the ending position for that position. If the tourist falls off the cliff, then that branch of the factor tree ends — you will not fill in every oval.

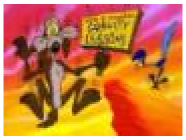
2) Compute the probability of moving toward safety or toward the edge for each step as shown in the diagram.

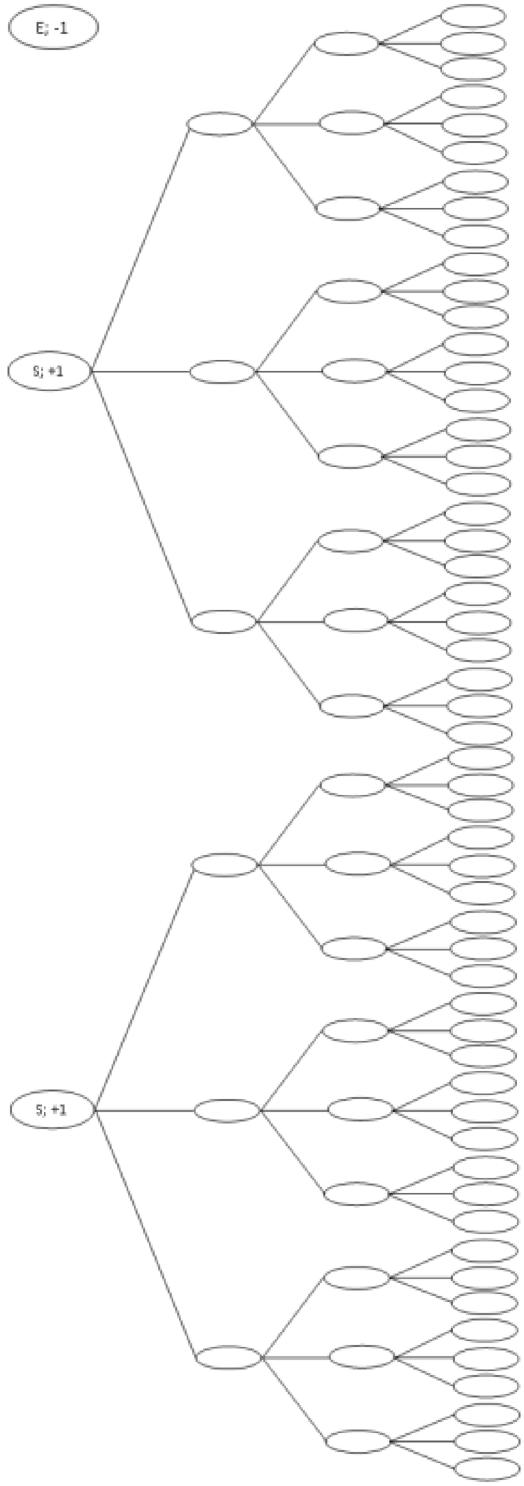
3) Based on your factor tree, what is the likelihood that the tourist will be safe?

4) Play five rounds of the cliff hanger game. Record the results below.

Game	Outcome (Win or Lose)
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

5) Do the repeated trials for the cliff hanger represent random trials? Why or why not?





Record the probability of moving toward safety or toward the edge with each step.
Note: These are conditional probabilities — each step is contingent on not having already fallen!

Step 1: P(Safety) =	If traveler takes Step 2: P(Safety) =	If traveler takes Step 3: P(Safety) = ________	If traveler takes Step 4: P(Safety) = _______

APPENDIX E

Algebra Lesson 3/Geometry Lesson 1: Counting Principles

Situation:

This lesson is the third of a probability unit designed for high school algebra classes. In this lesson, students will examine number properties important to probability. In the opener, students compare rational number quantities. In the main lesson, students learn to count possible outcomes using the Fundamental Counting Principle and factorial structures. Prior to this lesson, students have examined statistical structures, normal distribution, and notions of randomness. This unit will continue with exploration of probability and probability distributions.

This lesson is designed for a 90 minute block period class.

Objectives

4) Students will count outcomes using the Fundamental Counting Principle.
5) Students will count outcomes using factor trees.
6) Students will count outcomes using permutations.
7) Students will count outcomes using combinations.

Connections

Students have completed their first algebra unit. The opener for this lesson connects algebra, geometry, and rational numbers (meaning of rational numbers linked to central angle measurement and numerical notations). The main lesson lays a foundation for the rest of the probability unit.

Materials

12) LCD Projector
13) Microsoft PowerPoint (And Clicker, if available)
14) PowerPoint Presentation
15) Student Lesson Worksheet
16) Student Practice Worksheet
17) Student Practice Worksheet Key

KY Core Content 4.1 Standards

MA-HS-1.1.1: Students will compare real numbers using order relations (less than, greater than, equal to) and represent problems using real numbers.

MA-HS-4.4.2: Students will recognize and identify the differences between combinations and permutations and use them to count discrete quantities.

Procedures (88 minutes overall)

10) Opener – Think, Pair, Share (@ 8 minutes overall)
    a. Give @ 2 minutes to complete the work. *(Teacher takes roll and posts)*

b. Pair up and discuss answers (@ 2 minutes)
c. Share out in pairs (@ 2 minutes)
d. Discuss sample answers on PowerPoint slide. (@ 2 minute)
    i. Illustration connects angle degrees to rational number comparison
    ii. Before showing the number line, discuss why 99 is the smallest common denominator and what common denominators mean.

11) Pass out Student Lesson Worksheet. Have students read about Counting Principles in their textbook and answer as many as they can on their own. (5 – 10 minutes)
12) Move students into groups to discuss their answers. (5 minutes)
13) Students return to their normal seats for whole class discussion. (30 minutes)
    a. Events
        i. Independent Events
        ii. Dependent Events
    b. Counting Independent Outcomes: Three tosses of a coin
        i. <u>Factor Trees:</u>
            1. **Why are some of the outcomes in red?** *Because these flips would not be necessary to determine a winner in a "Best 2 out of 3" game.*
            2. **How many outcomes to 3 tosses of a coin? How do you know?** *8, They can be counted on the last row of the factor tree.*
        ii. <u>Table Arrangement:</u>
            1. **How do these outcomes match the factor tree?** *Each column on the table matches a level of the factor tree.*
            2. **Which do you think is easier to read? Why?** *Answers will vary.*
            3. **What does it mean to be systematic? Why is it important?** *Counting/Arranging in a pre-determined order; important to ensure all outcomes are counted.*
        iii. Fundamental Counting Principle
            1. Definition
            2. Discuss Notation from slide (| | = magnitude; $\cap$ = intersection). **Why multiply?** *Each outcome for the second event applies to every outcome for the first event.*
               When would you not multiply? *When a situation calls for the union of two sets instead of the intersection: Each outcome for B does not apply to every outcome for A.*
            3. **Why do you suppose the Fundamental Counting Principle extends to more than 2 events?** *Each subsequent outcome applies to every outcome for each previous event.*
            4. **How do we count 3 flips of a coin?** *2 outcomes for the first flip, 2 outcomes for the second flip, and 2 outcomes for the third flip = 2•2•2 = 8 outcomes*
            5. **Disadvantage?** *Doesn't list the outcomes as well as count them; however, more useful for large number of outcomes*

(e.g., roll 3 dice: 6•6•6 = 216 outcome, impractical to list them all)

c. Counting Dependent Events
   - **i. A classroom of 30 students arrives on the first day.**
   - **ii. Since I don't know any of the students I assign seats randomly.**
   - iii. Count the number of outcomes for each of the first five students: 30•29•28•27•26…
   - iv. At this point, ask the students to describe the pattern. $n!$ or 30!

d. Permutations
   - i. To introduce permutations, consider the following situation with the students: "The office calls and needs me to choose 3 people to help with a project." How many ways to choose? *30•29•28*
   - ii. Developing a formula:
     1. All possibilities = 30!
     2. To count only the first 3, we have to remove 27!
     3. To remove these numbers mathematically, divide n! by (n-r)!
   - **iii. Permutations are used when each new arrangement should be counted as a different outcome, or we say, "Order Matters."**
   - iv. Look at example of choosing a President, Vice President, and Secretary. **Why does order matter in this situation?**
   - v. Show how to compute permutation on graphing calculator.

e. Combinations
   - i. New situation: team of 3 instead of 3 different positions.
   - ii. Why doesn't order matter in this situation? Because every arrangement of 3 people is now the same team: ABC, ACB, BAC, BCA, CAB, CBA
   - iii. Show that 6 outcomes = 3!
   - iv. Levels of restrictiveness (Most outcomes to the least): Factorials → Permutations → Combinations
   - **v.** Develop combination formula from permutation formula: **To eliminate r! arrangements, have to divide. Why?** *Multiplicative Structure.*
   - **vi.** Discuss calculator functions and answer to example: number of teams of 3 out of 9 people.
   - **vii.** Discuss notation: ${}_nC_r = \binom{n}{r}$
   - **viii.** Have students calculate $\binom{4}{0}\binom{4}{1}\binom{4}{2}\binom{4}{3}\binom{4}{4}$
   - **ix.** Why do $\binom{n}{0}$ and $\binom{n}{n}$ always = 1?
   - **x.** Show Pascal's triangle. The r^{th} term of the n^{th} row = nCr. Emphasize that both must be counted from 0.

xi. Example 1: Twelve skiers compete in the final round of the Olympic freestyle skiing competition. How many different top three winners are possible? (Gold, Silver, Bronze).

1. **What are the events?** *A skier finishing.*
2. **Are these events independent or dependent?** *Dependent*
3. **Should the number of possible outcomes be counted with permutations or combinations? How do you know?** *Permutations because ABC is different than BCA.*
4. **Solve:** ${}_{12}P_3 = 1320$

xii. Example 1: Twelve skiers compete in the final round of the Olympic freestyle skiing competition. How many different top three winners are possible? (Gold, Silver, Bronze).

1. **What are the events?** *A skier finishing.*
2. **Are these events independent or dependent?** *Dependent*
3. **Should the number of possible outcomes be counted with permutations or combinations? How do you know?** *Permutations because ABC is different than BCA.*
4. **Solve:** ${}_{12}P_3 = 1320$

Omelets $7.95
(Each ingredient below adds an additional $0.50)

Vegetarian	Meat
Green Pepper	Ham
Red Pepper	Bacon
Onion	Sausage
Mushroom	Steak
Tomato	
Cheese	

Example 2: A restaurant serves omelets that can be ordered with any of the ingredients shown. (a) Suppose you want exactly 2 vegetarian ingredients and 1 meat ingredient in your omelet. How many different types of omelets can you order? (b) Suppose you can afford at most 3 ingredients in your omelet. How many different types of omelets can you order?

1. **What are the events?** *Ingredients chosen.*
2. **Are these events independent or dependent?** *Dependent*
3. **Should the number of possible outcomes be counted with permutations or combinations? How do you know?** *Combinations because Tomato and Cheese is the same as Cheese and Tomato.*
4. **Solve (a):** *Vegetarian:* ${}_6C_2 = 15$*; Meat:* ${}_4C_1 = 4$*; Vegetarian AND Meat =* $15 \cdot 4 = 60$
5. **Solve (b):** ${}_{10}C_3 = 120$

14) Pass out Student Practice Worksheet. Students work individually or with partners as desired. (15 minutes)
15) Discuss Counting Worksheet: Divide class into 3 groups. Each group discusses one question and presents their solution to the class. (5 minutes for group time; 15 minutes for reporting out)
16) Closure: 3-2-1
    a. What are 3 things that you found interesting today?
    b. What are 2 things that you learned?
    c. What is 1 thing you still have a question about?

Assessment

1) Opener: Students compare rational number values (1.1.1)
2) Guided Notes and Example Problems (Teacher observes and questions while students work individually; Questioning during whole class discussion) (4.4.2)
3) Counting Worksheet (4.4.2): Students work individually and collaboratively to analyze whether a situation involves independent or dependent events, Fundamental Counting Principle or Factorials, and permutations or combinations.

APPENDIX F

Algebra Lesson 3/Geometry Lesson 1 Task Rotation Prompts

1. You are going to set up a stereo system by purchasing separate components. In your price range you find 5 different receivers, 8 different compact disc players, and 12 different speaker systems.

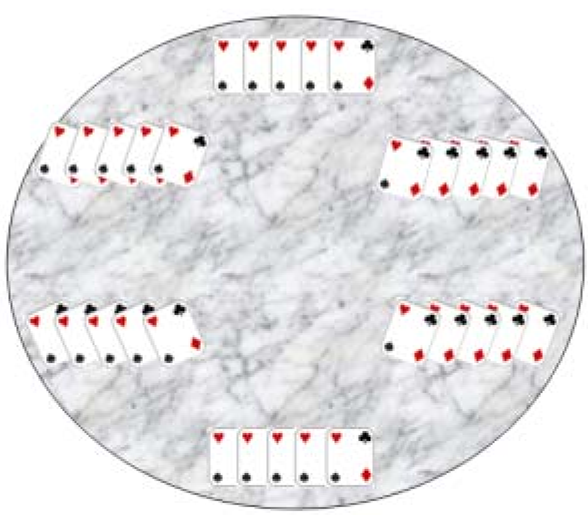
    a) What are the three events?

    b) Are these events independent or dependent? How did you decide?

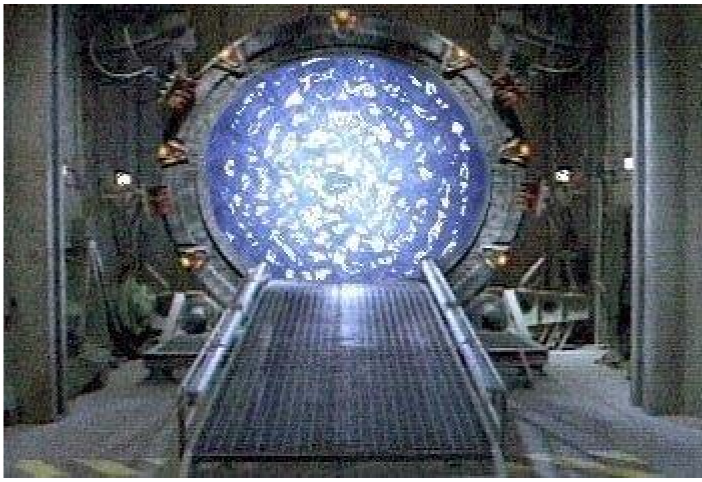
    c) If you want one of each of these components, how many different stereo systems are possible?

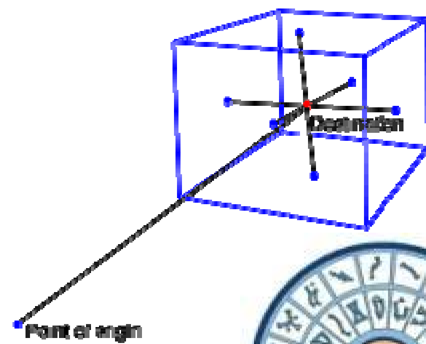
    d) Draw a picture to illustrate why multiplication is appropriate for counting the number of outcomes.

2. A deck of cards with no wilds is used for a hand of 5 card draw in a game with 6 players. Use the illustration below to demonstrate how to count the number of possible games that could be dealt.



3. In 1920, an Egyptologist discovered a StarGate, a means of travelling to planets all over the galaxy instantaneously. In 1994, Dr. Daniel Jackson discovered that the gate required seven symbols: Six points in space to identify the target planet and the point of origin.





The dialing device (DHD) has 39 symbols. Each time a planet is "dialed," seven symbols must be entered.

a) Does this situation suggest a permutation or combination? How can you tell?

b) How many planets could possibly be reached from a single DHD?

c) The point of origin is always the 7^{th} symbol. How does this information change your answer to (b)?

APPENDIX G

Algebra Lesson 4/Geometry Lesson 2: Event Probability

Situation:

This lesson is the fourth of a probability unit designed for high school algebra classes. In this lesson, students will rotate between stations to explore probability concepts. This lesson is designed for a 90 minute block period class.

Objectives

1) Students will differentiate between theoretical and experimental probability.
2) Students will evaluate problems using probability principles.
3) Students will explain how the law of large numbers applies to simulation.
4) Students will run a Monte Carlo simulation and interpret the outcome.
5) Students will use area and length ratios to compare probabilities.

Connections

In the previous lesson, students learned to count outcomes and differentiate between independent and dependent counting. This lesson extends the counting structures for each type of situation to probability.

Materials

18) Dice (For Stations 2 and 4)
19) LCD Projector
20) 2 Computers for Student Use (1 with Internet)
21) Microsoft PowerPoint (And Clicker, if available)
22) PowerPoint Presentation
23) Student Lesson Worksheet

KY Core Content 4.1 Standards

MA-HS-4.1.2: Students will construct data displays for data with no more than two variables.

MA-HS-4.4.1: Students will determine theoretical and experimental (from given data) probabilities, make predictions and draw inferences from probabilities, compare theoretical and experimental probabilities, and determine probabilities involving replacement and non-replacement.

MA-HS-4.4.3: Students will represent probabilities in multiple ways, such as fractions, decimals, percentages and geometric area models.

MA-HS-4.4.4: Students will explain how the law of large numbers can be applied in simple examples.

Procedures

0) Before class, Set up Stations:

a. **Station 1:** Computer with Applet: Spinners.

b. **Station 2**: Dice for Happy Meal Simulation
c. **Station 3**: Computer with Microsoft Excel; Print out directions for creating a Monte Carlo Simulation
d. **Station 4**: Dice for Craps Simulation

1) Opener (4 minutes overall)
    a. Give @ 2 minutes to complete the work. *(Teacher takes roll and posts)*
    b. Discuss sample answers on PowerPoint slide. (2 minutes)
2) Pass out Student Lesson Worksheet. Have students read about Probability in their textbook and answer as many as they can on their own. (5 minutes)
3) Move students into groups to discuss their answers. (5 minutes)
4) Students return to their normal seats for whole class discussion. (10 minutes)
    - Fundamental Concepts of Probability
        - "Probability" means the likelihood of an event occurring.
        - Expressed as a part-whole ratio
    - Success and Failure
        - Success
            - The outcome of interest.
            - Changes for each new situation.
        - Failure: Everything other than success
        - S = Number of Successful Outcomes
        - F = Number of Failure Outcomes
    - Theoretical Probability
        - Probability based on assumption that all outcomes are equally likely
        - Examples:
        - Tossing a Coin: P(H) = 0.5 = 50% = 1/2
        - Rolling a Single Die: P(3) = 0.167 = 16.7% = 1/6
        - Rolling 2 Dice: P(3) = 0.056 = 5.6% = 2/36 = 1/18
        - "3" from (1 and 2) or (2 and 1); 36 total outcomes
    - Experimental Probability
        - Probability Based on Observations, Data, or Simulation.
        - Examples from a random sample of 10 observations:
            - Coin Toss:
            - P(H) =
            - 1 Die:

        P(3) =
            - 2 Dice:

        P(3) =
    - Odds
        - Odds of Success: Ratio of Success : Failure
        - Odds of Failure: Ratio of Failure : Success
    - Example
        - A baseball player has 126 hits in 410 at-bats this season.
            - What is the probability that he gets a hit in his next at-bat?
            - Is this a theoretical or experimental probability?
            - What are his odds of success?

5) Station work (15 minutes each; 60 minutes total)

6) Closure: (5 minutes)
    a. You are the teacher in a class that just completed today's lesson.
    b. Write down three details you think are important for students to know from this lesson.
    c. Share out in class.

Assessment

1) Opener: Review of Fundamental Counting Principle
2) Student Worksheet (4.1.2, 4.4.1, 4.4.3, 4.4.4): Students work individually and collaboratively to explore fundamental probability concepts, probability ratios, and Monte Carlo simulations. Exploratory problems include multiple representations of probability (i.e., fractions, percentages, decimals).

APPENDIX H

Algebra Lesson 4 Station Prompts

Station 1: Spinners.

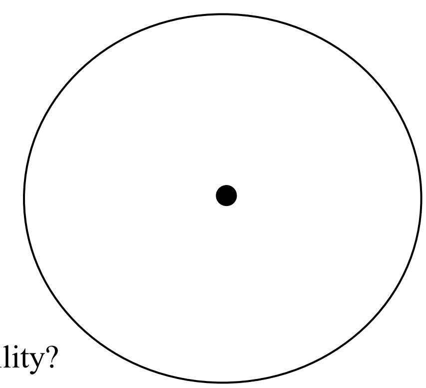
1) Refer to Applet: “Spinners.”
   a) What is the probability of the spinner landing on Purple? Green? Red? Orange? Yellow?

   b) Are these probabilities theoretical or experimental? How do you know?

   c) Click on the “Record Results” button. Record your data in the table below.
      i. Spin the dial 10 times.
      ii. Spin the dial 10 more times.
      iii. In the box that says “Spins,” type in “10.” Spin 8 more times.
      iv. In the box that says “Spins,” type in “50.” Spin 8 more times.

	Purple		Green		Red		Orange		Yellow	
	Number	Probability	Number	Probability	Number	Probability	Number	Probability	Number	Probability
10 spins										
20 spins										
100 spins										
500 spins										

   d) Click on the “Change Spinner” button. Change the values of each color as follows:
      Purple: 2
      Green: 3
      Red: 1
      Orange: 1
      Yellow: 4
   Click “Apply.”
   Draw a picture of the new spinner



      i. How are the new numbers related to probability?

      ii. Repeat the experiment above with your new spinner.

	Purple		Green		Red		Orange		Yellow	
	Number	Probability	Number	Probability	Number	Probability	Number	Probability	Number	Probability
10 spins										
20 spins										
100 spins										
500 spins										

   e) How do your two experiments demonstrate the Law of Large Numbers?

Station 2: Happy Meal Simulation

2) A blogger posted the following comment about the McDonald's Happy Meal prize:

   My daughter loves the current McDonald's happy meal prize. It's a Kids Bop CD, and it rocks! There are 6 different ones, and they are definitely doing the job, as we are pursuing to collect all 6! I feel like it's really worth it. Some of the prizes end up in the trash (when the kids aren't looking), but these will be around for a while. I am a fan of anything that makes the car trips easier:).

   Suppose that the prizes are randomly placed in bags. Run a simulation to determine how many Happy Meals are likely to need to be bought to get all 6 Kids Bop CD's. Let each roll of a die represent the purchase of a different Kids Bop CD.

   a) Roll the dice until you roll all 6 numbers. Use tick marks to record the outcome for each roll in the table below.

Outcome	**CD 1**	**CD 2**	**CD 3**	**CD 4**	**CD 5**	**CD 6**
No. of Rolls						
Probability Ratio						
Percentage						

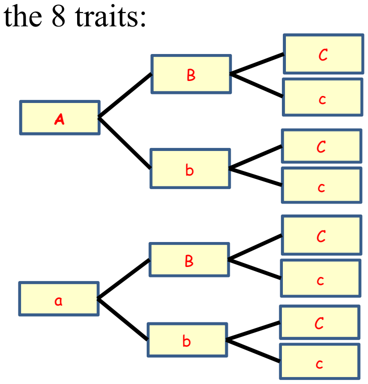
   b) How many rolls did it take you to buy all 6 CD's?

c) What does the law of large numbers indicate about the relationship of your experimental probability to the theoretical probability?

Station 3: Monte Carlo Simulation

3) What is a Monte Carlo simulation?

4) In Meiosis, chromosomes from father and mother join to create a new gene. A scientist is studying 3 genes, each with a dominant and recessive trait. Using a factor tree, she determined that 8 traits were possible: A cereal company is putting 8 different prizes in their boxes. Run a Monte Carlo simulation to determine how many boxes you'll have to buy to get all 8 prizes.
   a) Double Click the Microsoft Excel Template on the desktop, "G2_Station_3_Monte_Carlo_Simulation.xltx"
   b) Use the factor tree on the "Introduction" page to determine the gene labels for the 8 traits:



Trait 1: _______
Trait 2: _______
Trait 3: _______
Trait 4: _______
Trait 5: _______
Trait 6: _______
Trait 7: _______
Trait 8: _______

   c) What is the theoretical probability for each trait?
   d) Go to the "Monte Carlo" page. Type in "=RandBetween(1,8)" into cell A2 and hit Enter.
   e) Click and hold the button at the bottom right corner of cell A2 as shown to the left; drag the pointer down to A10.

      i. What do you suppose each cell represents?
      ii. Has the scientist encountered all 8 traits yet? How can you tell?
      iii. Does the experimental probability represent the theoretical probability? How do you know?

   f) Click and hold the button at the bottom right corner of cell A10. Drag to cell A25.
      i. Has the scientist encountered all 8 traits yet?
      ii. Does the experimental probability represent the theoretical probability? How do you know?

g) Click and hold the button at the bottom right corner of cell A25. Drag to cell A150. Does the experimental probability represent the theoretical probability now? How can you tell?

h) Click and hold the button at the bottom right corner of cell A150. Drag to cell A350. Does the experimental probability represent the theoretical probability now? How can you tell?

i) Click and hold the button at the bottom right corner of cell A350. Drag to cell A500. Does the experimental probability represent the theoretical probability now? How can you tell?

j) Click and hold the button at the bottom right corner of cell A500. Drag to cell A1000. Does the experimental probability represent the theoretical probability now? How can you tell?

k) Based on this simulation, what sample size is needed to ensure that the sample data will represent the population distribution? How did you decide?

l) Explain the relationship of the Histograms A, B, C, and D to the simulation data? To each other?

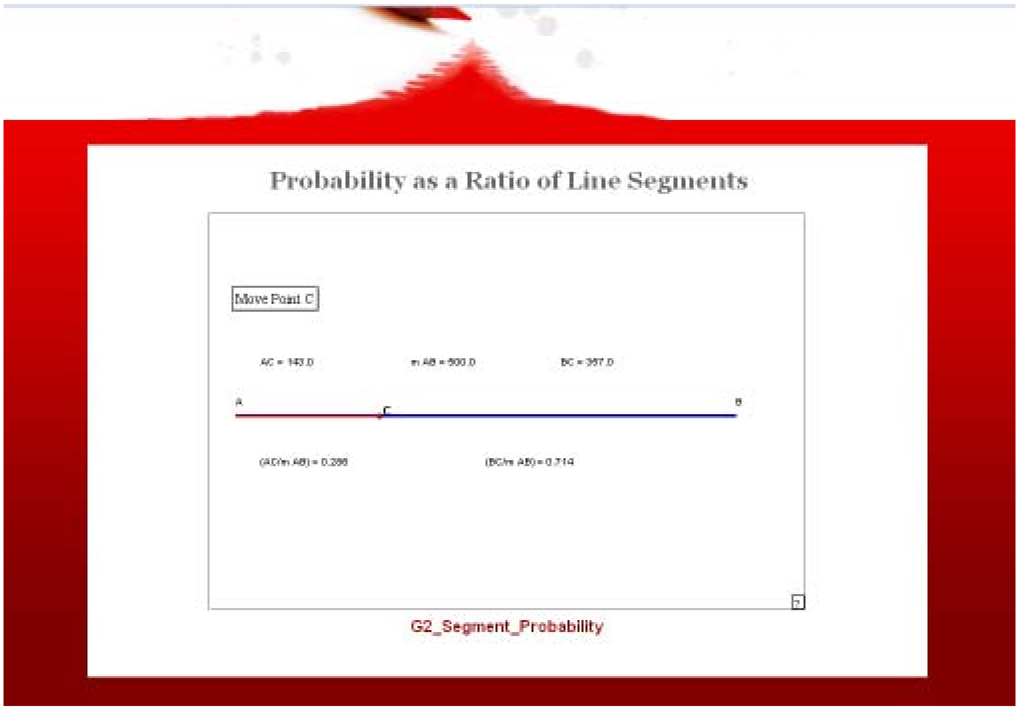
m) What are 2 advantages of a Monte Carlo simulation over other simulations (e.g., dice, spinners, coins)? 2 disadvantages?

APPENDIX I

Geometry Lesson 2 Station Prompts

Station 1: Geometric Probability.

1) Refer to Applet: "Probability as a ratio of line segments."

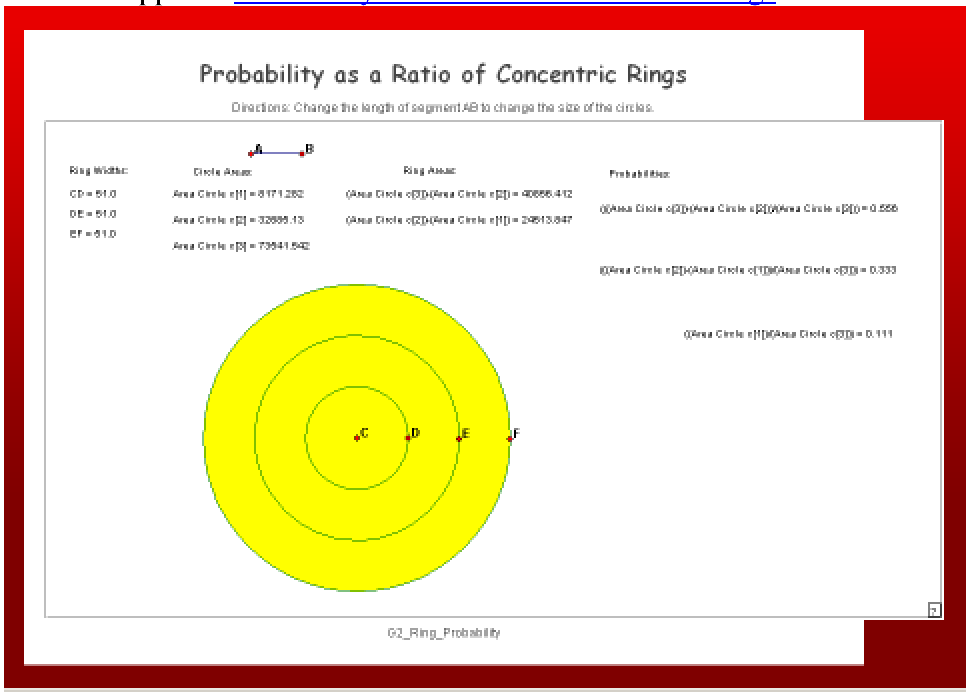


a) If segment AB represents the entire set outcome, what is the probability of AC, P(AC)?

b) What is the relationship of P(BC) to P(AC)? How does this relationship relate to the Segment Addition Property?

c) What is the probability that $P(BC) \geq 0.80$? How did you decide?

2) Refer to Applet: “Probability as a Ratio of Concentric Rings.”



a) What is the relationship of the radii of the three rings?

b) Which measurement represents the area of the outer ring? Why?

c) Which measurement represents the area of the middle ring? Why?

d) Which measurement represents the area of the bulls eye? How is this region different from the other two?

e) What is the probability of a randomly thrown dart landing in the outer ring? How do you know?

f) What is the probability of a randomly thrown dart landing in the middle ring? How do you know?

g) What is the probability of a randomly thrown dart landing in the bulls eye? How do you know?

h) Why don’t the probability measurements change as the circle moves?

i) Why aren’t the probabilities of each region equal?

Station 2: Happy Meal Simulation

3) A blogger posted the following comment about the McDonald's Happy Meal prize:

   My daughter loves the current McDonald's happy meal prize. It's a Kids Bop CD, and it rocks! There are 6 different ones, and they are definitely doing the job, as we are pursuing to collect all 6! I feel like it's really worth it. Some of the prizes end up in the trash (when the kids aren't looking), but these will be around for a while. I am a fan of anything that makes the car trips easier:).

   Suppose that the prizes are randomly placed in bags. Run a simulation to determine how many Happy Meals are likely to need to be bought to get all 6 Kids Bop CD's. Let each roll of a die represent the purchase of a different Kids Bop CD.

   d) Roll the dice until you roll all 6 numbers. Use tick marks to record the outcome for each roll in the table below.

Outcome	**CD 1**	**CD 2**	**CD 3**	**CD 4**	**CD 5**	**CD 6**
No. of Rolls						
Probability Ratio						
Percentage						

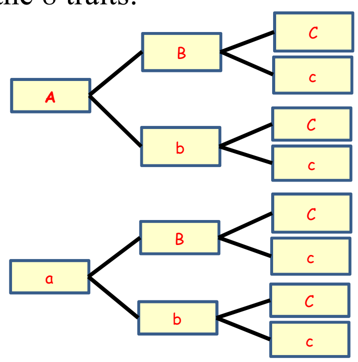
   e) How many rolls did it take you to buy all 6 CD's?

   f) What does the law of large numbers indicate about the relationship of your experimental probability to the theoretical probability?

Station 3: Monte Carlo Simulation

4) What is a Monte Carlo simulation?

5) In Meiosis, chromosomes from father and mother join to create a new gene. A scientist is studying 3 genes, each with a dominant and recessive trait. Using a factor tree, she determined that 8 traits were possible: A cereal company is putting 8 different prizes in their boxes. Run a Monte Carlo simulation to determine how many boxes you'll have to buy to get all 8 prizes.
   a) Double Click the Microsoft Excel Template on the desktop, "G2_Station_3_Monte_Carlo_Simulation.xltx"
   b) Use the factor tree on the "Introduction" page to determine the gene labels for the 8 traits:



Trait 1: ________
Trait 2: ________
Trait 3: ________
Trait 4: ________
Trait 5: ________
Trait 6: ________
Trait 7: ________
Trait 8: ________

   c) What is the theoretical probability for each trait?
   d) Go to the "Monte Carlo" page. Type in "=RandBetween(1,8)" into cell A2 and hit Enter.
   e) Click and hold the button at the bottom right corner of cell A2 as shown to the left; drag the pointer down to A10.

      i. What do you suppose each cell represents?
      ii. Has the scientist encountered all 8 traits yet? How can you tell?

      iii. Does the experimental probability represent the theoretical probability? How do you know?

   f) Click and hold the button at the bottom right corner of cell A10. Drag to cell A25.
      i. Has the scientist encountered all 8 traits yet?

      ii. Does the experimental probability represent the theoretical probability? How do you know?

g) Click and hold the button at the bottom right corner of cell A25. Drag to cell A150. Does the experimental probability represent the theoretical probability now? How can you tell?

h) Click and hold the button at the bottom right corner of cell A150. Drag to cell A350. Does the experimental probability represent the theoretical probability now? How can you tell?

i) Click and hold the button at the bottom right corner of cell A350. Drag to cell A500. Does the experimental probability represent the theoretical probability now? How can you tell?

j) Click and hold the button at the bottom right corner of cell A500. Drag to cell A1000. Does the experimental probability represent the theoretical probability now? How can you tell?

k) Based on this simulation, what sample size is needed to ensure that the sample data will represent the population distribution? How did you decide?

l) Explain the relationship of the Histograms A, B, C, and D to the simulation data? To each other?

m) What are 2 advantages of a Monte Carlo simulation over other simulations (e.g., dice, spinners, coins)? 2 disadvantages?

Station 4: Craps

In the game of craps, two common bets are pass line bets and don't pass line bets.

Pass Line Bet - You win if the first roll is a natural (7, 11) and lose if it is craps (2, 3, 12). If a point is rolled (4, 5, 6, 8, 9, 10) it must be repeated before a 7 is **thrown** in order to win. If 7 is **rolled** before the point you lose.

Don't Pass Line Bet - This is the reversed Pass Line bet. If the first roll of a dice is a natural (7, 11) you lose and if it is a 2 or a 3 you win. A dice roll of 12 means you have a tie or push with the casino. If the roll is a point (4, 5, 6, 8, 9, 10) a 7 must come out before that point is repeated to make you a **winner**. If the point is rolled again before the 7 you lose.

1) Develop the probability distribution for rolling two die.
   a) List all possible dice sums in the table below.

Die 1	Die 2	Outcome	Die 1	Die 2	Outcome	Die 1	Die 2	Outcome
1	1		3	1		5	1	
	2			2			2	
	3			3			3	
	4			4			4	
	5			5			5	
	6			6			6	
2	1		4	1		6	1	
	2			2			2	
	3			3			3	
	4			4			4	
	5			5			5	
	6			6			6	

   b) What is the total number of outcomes for rolling two die?
   c) Record the probability of each outcome for rolling two dice.

Sum	2	3	4	5	6	7	8	9	10	11	12
Probability											

   d) Create a histogram below of the data in the table.

2) What does this distribution tell you about the most likely outcomes for rolling two dice?

3) Would you prefer to bet on a Pass Bet or No Pass Bet? How did you decide?

4) Play 10 rounds of craps, using the type of bet you chose. Record the results below.

Roll	Outcome	Win or Lose?
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

5) The Law of large numbers states that as sample size increases, its ability to represent the population also increases. How do you think the law of large number relate to your simulation of craps?

APPENDIX J

Algebra Lesson 5/Geometry Lesson 3: Probability Distributions

Situation:

This lesson is the final lesson of a probability unit designed for high school algebra and geometry classes.
This lesson is designed for a 90 minute block period class.

Objectives

1) Students will compare two different experiments with equal probability using the Law of Large Numbers.
2) Students will make inferences about populations using binomial and geometric distribution patterns.

Connections

In the previous lesson, students examined probability as part-whole relationships and single and multiple events. They also explored the impact of the law of large numbers on the relationship between theoretical and experimental probability. In this lesson, students will examine probability distributions as an extension of event probabilities.

Materials

1) Microsoft PowerPoint (And Clicker, if available)
2) PowerPoint Presentation
3) Student Lesson Worksheet
4) Task Rotation Worksheet

KY Core Content 4.1 Standards

MA-HS-4.1.1: Students will analyze and make inferences from a set of data with no more than two variables, and will analyze problems for the use and misuse of data representations.

MA-HS-4.1.2: Students will construct data displays for data with no more than two variables.

MA-HS-4.2.1: Students will describe and compare data distributions and make inferences from the data based on the shapes of graphs, measures of center (mean, median, mode) and measures of spread (range, standard deviation).

MA-HS-4.4.4: Students will explain how the law of large numbers can be applied in simple examples.

MA-HS-5.1.3: Students will demonstrate how equations and graphs are models of the relationship between two real-world quantities (e.g., the relationship between degrees Celsius and degrees Fahrenheit)

Procedures

1) Opener (6 minutes): The theoretical probability for an event is ½. Which of the two ratios is more likely in a set of repeated trials? 4/8 or 400/800.
    a. Give students @ 2 minutes to complete. (*Teacher takes roll and posts*)

b. In pairs, have students share their thoughts about the ratio likelihoods. (2 minutes)
c. Report out and discuss with whole class (The slide "Law of Large Numbers" is a discussion of opener solution; 2 minutes)

2) Introduce Probability Distributions (28 minutes total)
    a. Fundamental Ideas (3 minutes)
        i. Definition
        ii. Two types: Discrete or Continuous
        iii. Based on Random Data and Patterns that emerge from repeated trials.
    b. Uniform Distribution (5 minutes)
        i. Uses
        ii. Examples
        iii. Shape of the distribution
        iv. Interpretation
    c. Binomial Distribution (10 minutes)
        i. Uses
        ii. Examples
        iii. Shape of the distribution
        iv. Interpretation
        v. Example Problem
    d. Geometric Distribution (10 minutes)
        i. Uses
        ii. Examples
        iii. Shape of the distribution
        iv. Interpretation
        v. Example problem

3) Problem Set (50 minutes). *Pass out Student Problem Set Worksheet.*
    a. Do each question one at a time.
    b. Allow students to work approximately 10 minutes individually.
    c. Have students get with a partner and discuss (10 minutes). Use different partners for each question.
    d. Report out and discuss with whole class. (5 minutes).
    e. Repeat for 2nd question.
4) Closure (5 minutes)
    a. Give a couple of minutes for individual work.
    b. Report out with whole class.

Assessment

1) Opener assesses how well students can apply Law of Large Numbers to various situations (4.1.1, 4.4.4)
2) Problem Set Question 1 (Blood Type) assesses how well students can apply binomial distribution inference to a particular situation. (4.1.2, 4.2.1)
3) Task Rotation Question 2 (Daughter) assesses how well students can apply the geometric distribution to a particular situation. Additionally, students are asked to interpret the probability histogram. (4.1.2, 4.2.1, 5.1.3)

APPENDIX K

Algebra Lesson 5 Student Worksheet

1. Blood type is inherited. Suppose a father carries the genotype AO (phenotype = Type A) while the mother carries the genotype BO (phenotype = type B). They have 4 children.

   a) Fill in the Punnett Square:

Mother: \ Father:		

Possible Genotypes	Probability

Possible Phenotypes	Probability

   b) Type O blood is considered the "Universal Donor." What is the probability that X = 2 of the children will have Type O blood?

      i. Why is this situation binomial?

      ii. What does X represent?

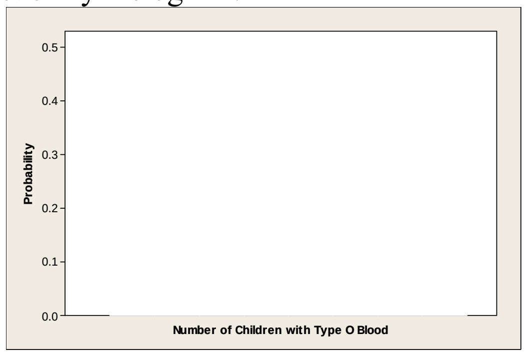
      iii. n = ?

      iv. p = ?

      v. Use the graphing calculator to calculate the full probability distribution of X. Then calculate the probability distribution of X on Minitab.

Number of Type O Children:	X = ?	X = ?	X = ?	X = ?	X = ?	**Sum of Probabilities**
Probability						

c) Draw the probability histogram:



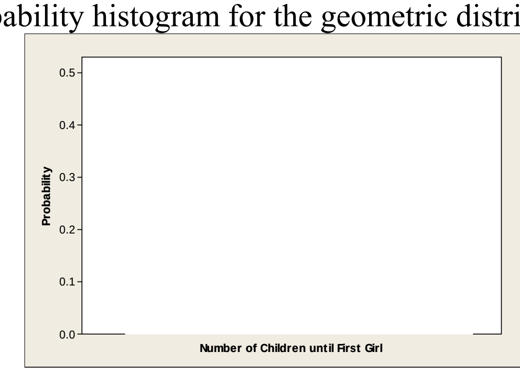
d) What is the probability that 2 children from this family will have Type O Blood?

2. A couple decided to start a family, and both wanted a daughter. Their first child was a boy, so they decided to continue having children until they got a daughter. After having four boys, the couple's fifth child was a daughter.

   a) In what way is this situation geometric?

   b) After the fourth child's gender was known, the mother proclaimed, "What are the chances?!" What was the probability that it would take 5 children before a girl was born?

   c) Draw a probability histogram for the geometric distribution out to 5 children.



   d) Does this graph mean that it is less likely that the 2nd child is a girl than the 1st? Why or why not? If not, what does it mean?

   e) Develop an equation to represent the geometric probability of achieving the first success on the fifth try.

APPENDIX L

Geometry Lesson 3 Student Worksheet

The useful life of a radial tire is normally distributed with a mean of 80,000 miles and a standard deviation of 5000 miles. The company makes 10,000 tires a month.

a. **About how many tires from a month's production will last between 75,000 and 85,000 miles?**
b. **About how many tires from a month's production will last more than 90,000 miles?**
c. **What is the probability that if you buy a radial tire at random, it will last between 70,000 and 85,000 miles?**
d. **As a consumer, what are two things you can do to maximize the life of your tires?**

Mario and Luigi are calculating the probability of getting a 4 and then a 2 if they roll a die twice. With a partner, decide which solution is correct and why.

Mario:
P(4, then 2) Independent Events, so the probability is:

$$\frac{1}{6} \bullet \frac{1}{6}$$

$$= \frac{1}{36}$$

= 2.78% chance of rolling a 4, then a 2.

Luigi:
P(4, then 2) Dependent Events, so the probability is:

$$\frac{1}{6} \bullet \frac{1}{5}$$

$$= \frac{1}{30}$$

= 3.33% chance of rolling a 4, then a 2.

In 1998, Ben took both the SAT and the ACT. On the mathematics section of the SAT, he earned a score of 624. On the mathematics section of the ACTG, he earned a score of 31.

For the SAT, the mean was 512 and the standard deviation was 112. For the ACT, the mean was 21 and the standard deviation was 5.

- **Explain how you can know that Ben performed better on the ACT than he did on the SAT.**

Sketch three normal curves on the same scale with the following properties (you can use the graphing calculator):

a. **Mean is 50 and standard deviation is 2.**
b. **Mean is 50 and standard deviation is 10.**
c. **Mean is 50 and standard deviation is 20.**

If you owned a business and the data represents your profits each month over the past year, which curve would you prefer? Why?

APPENDIX M

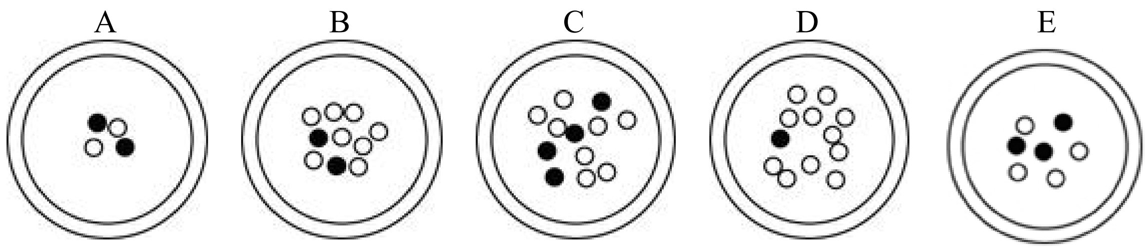
NAEP Mathematics Knowledge Instrument

I. Multiple Choice Answers	Explanations
1) A B C D E	
2) A B C D E	
3) A B	
4) A B C D E	
5) A B C D E	
6) A B C D E	
7) A B C D E	
8) A B C D E	
9) A B	
10) A B C D	
11) A B C D E	

12) A B C D E	
13) A B C D E	
14) A B C D E	
15) A B C D E	
16) A B C D E	
17) A B C D E	

I. Multiple Choice. For Questions 1-15, please mark the response that you think best answers the question. Please explain how you decided your answer.

1) A person is going to pick one marble without looking. For which dish is there the greatest probability of picking a black marble?



Please explain your answer.

2) The table below shows the gender and color of 7 puppies. If a puppy selected at random from the group is brown, what is the probability it is a male?

	Male	Female
Black	1	2
Brown	1	3

A) $\frac{1}{4}$ B) $\frac{2}{7}$ C) $\frac{1}{3}$ D) $\frac{1}{2}$ E) $\frac{2}{3}$

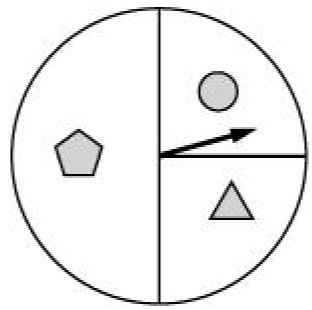
Please explain your answer.

3) A package of candies contained only 10 red candies, 10 blue candies, and 10 green candies. Bill shook up the package, opened it, and started taking out one candy at a time and eating it. The first 2 candies he took out and ate were blue. Bill thinks the probability of getting a blue candy on his third try is $\frac{10}{30}$ or $\frac{1}{3}$.

a) Is Bill correct or incorrect? A) Yes, he is correct. B) No, he is not correct.

b) Please explain how you decided.

4) If Rose spins a spinner like the one below 300 times, about how many times should she expect it to land on the space with a circle?



A) 75 B) 90 C) 100 D) 120 E) 150

Please explain how you decided.

5) The temperature in degrees Celsius can be found by subtracting 32 from the temperature in degrees Fahrenheit and multiplying the result by $\frac{5}{9}$. If the temperature of a furnace is 393 degrees Fahrenheit, what is it in degrees Celsius, to the nearest degree?

A) 650 B) 1805 C) 40 D) 201 E) 72

Please explain how you decided.

6) In the equation $y = 4x$, if the value of x is increased by 2, what is the effect on the value of y?

A) It is 8 more than the original amount.
B) It is 6 more than the original amount.
C) It is 2 more than the original amount.
D) It is 16 times the original amount.
E) It is 8 times the original amount.

Please explain how you decided.

7) A plumber charges customers $48 for each hour worked plus an additional $9 for travel. If h represents the number of hours worked, which of the following expressions could be used to calculate the plumber's total charge in dollars?

A) $h + 48 + 9$
B) $48 \bullet 9 \bullet h$
C) $9h + 48$
D) $h + (48 \bullet 9)$
E) $48h + 9$

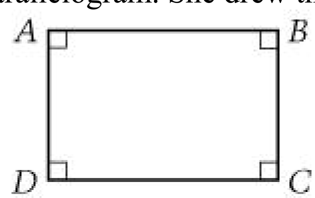
Please explain how you decided.

8) At the school carnival, Carmen sold 3 times as many hot dogs as Shawn. The two of them sold 152 hot dogs altogether. How many hot dogs did Carmen sell?

A) 21 B) 38 C) 51 D)114 E) 148

Please explain how you decided.

9) Sara was asked to draw a parallelogram. She drew the figure below.

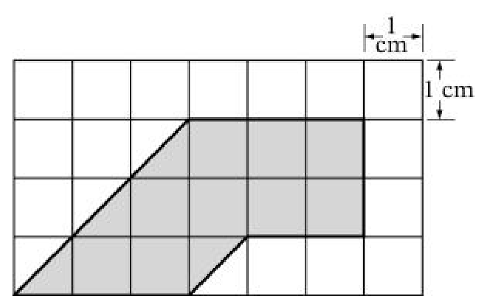


Is Sara's figure a parallelogram?

A) Yes B) No

Please explain how you decided.

10) What is the area of the shaded figure?



A) 9 square centimeters
B) 11 square centimeters
C) 13 square centimeters
D) 14 square centimeters

Please explain how you decided.

11) A scale drawing of a rectangular room is 5 inches by 3 inches. If 1 inch on this scale drawing represents 3 feet, what are the dimensions of the room?

A) 5 feet by 3 feet
B) 5 feet by 9 feet
C) 15 feet by 3 feet
D) 15 feet by 5 feet
E) 15 feet by 9 feet

Please explain how you decided.

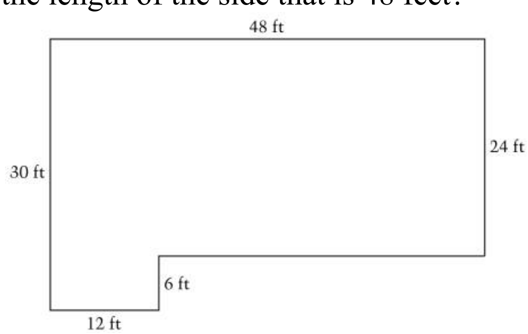
12) The cost to mail a first-class letter is 33 cents for the first ounce. Each additional ounce costs 22 cents. (Fractions of an ounce are rounded up to the next whole ounce.)

How much would it cost to mail a letter that weighs 2.7 ounces?

A) 55 cents B) 66 cents C) 77 cents D) 88 cents E) 99 cents

Please explain how you decided.

13) If you were to redraw the diagram using a scale of $\frac{3}{4}$ inch = 10 feet, what would be the length of the side that is 48 feet?



A) 3.0 in B) 3.6 in C) 5.6 in D) 7.5 in E) 12.0 in

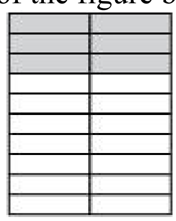
Please explain how you decided.

14) In which of the following are the three fractions arranged from least to greatest?

A) $\frac{2}{7}, \frac{1}{2}, \frac{5}{9}$ B) $\frac{1}{2}, \frac{2}{7}, \frac{5}{9}$ C) $\frac{1}{2}, \frac{5}{9}, \frac{2}{7}$ D) $\frac{5}{9}, \frac{1}{2}, \frac{2}{7}$ E) $\frac{5}{9}, \frac{2}{7}, \frac{1}{2}$

Please explain how you decided.

15) What fraction of the figure below is shaded?



A) $\frac{1}{4}$ B) $\frac{3}{10}$ C) $\frac{1}{3}$ D) $\frac{3}{7}$ E) $\frac{7}{10}$

Please explain how you decided.

16) Angela makes and sells special-occasion greeting cards. The table below shows the relationship between the number of cards sold and her profit. Based on the data in the table, which of the following equations shows how the number of cards sold and profit (in dollars) are related?

	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.
Number Sold, n	4	0	5	2	3	6
Profit, p	\$2.00	\$0.00	\$2.50	\$1.00	\$1.50	\$3.00

A) $p = 2n$ B) $p = 0.5n$ C) $p = n - 2$ D) $p = 6 - n$ E) $p = n + 1$

Please explain how you decided.

17) Each of the 6 faces of a certain cube is labeled either R or S. When the cube is tossed, the probability of the cube landing with an R face up is $\frac{1}{3}$.How many faces are labeled R?

A) Five B) Four C) Three D) Two E) One

APPENDIX N

Attitudes Toward Mathematics Inventory (Tapia & Marsh, 2004)

This inventory consists of statements about your attitude toward mathematics. There are no correct or incorrect responses. Read each item carefully. Please think about how you feel about each item. Circle the letter that most closely corresponds to how the statements best describes your feelings. Use the following response scale to respond to each item. Your responses are confidential.	Use these codes: A-Strongly Disagree B-Disagree C-Neutral D-Agree E-Strongly Agree
1) Mathematics is a very worthwhile and necessary subject.	A B C D E
2) I want to develop my mathematical skills.	A B C D E
3) I get a great deal of satisfaction out of solving a mathematics problem.	A B C D E
4) Mathematics helps develop the mind and teaches a person to think.	A B C D E
5) Mathematics is important in everyday life.	A B C D E
6) Mathematics is one of the most important subjects for people to study.	A B C D E
7) High school math courses would be very helpful no matter what I decide to study.	A B C D E
8) I can think of many ways that I use math outside of school.	A B C D E
9) Mathematics is one of my most dreaded subjects.	A B C D E
10) My mind goes blank and I am unable to think clearly when working with mathematics.	A B C D E
11) Studying mathematics makes me feel nervous.	A B C D E
12) Mathematics makes me feel uncomfortable.	A B C D E
13) I am always under a terrible strain in a math class.	A B C D E
14) When I hear the word mathematics, I have a feeling of dislike.	A B C D E
15) It makes me nervous to even think about having to do a mathematics problem.	A B C D E
16) Mathematics does not scare me at all.	A B C D E
17) I have a lot of self-confidence when it comes to mathematics	A B C D E
18) I am able to solve mathematics problems without too much difficulty	A B C D E
19) I expect to do fairly well in any math class I take.	A B C D E
20) I am always confused in my mathematics class.	A B C D E
21) I feel a sense of insecurity when attempting mathematics.	A B C D E
22) I learn mathematics easily.	A B C D E
23) I am confident that I could learn advanced mathematics.	A B C D E
24) I have usually enjoyed studying mathematics in school.	A B C D E
25) Mathematics is dull and boring.	A B C D E
26) I like to solve new problems in mathematics	A B C D E
27) I would prefer to do an assignment in math than to write an essay.	A B C D E
28) I would like to avoid using mathematics in college.	A B C D E
29) I really like mathematics.	A B C D E
30) I am happier in a math class than in any other class.	A B C D E
31) Mathematics is a very interesting subject.	A B C D E
32) I am willing to take more than the required amount of mathematics.	A B C D E
33) I plan to take as much mathematics as I can during my education.	A B C D E
34) The challenge of math appeals to me.	A B C D E
35) I think studying advanced mathematics is useful.	A B C D E
36) I believe studying math helps me with problem solving in other areas.	A B C D E
37) I am comfortable expressing my own ideas on how to look for solutions to a difficult problem in math.	A B C D E
38) I am comfortable answering questions in math class.	A B C D E
39) A strong math background could help me in my professional life.	A B C D E
40) I believe I am good at solving math problems.	A B C D E

APPENDIX O

Metacognitive Awareness Inventory (Schraw & Dennison, 1994)

We would like you to respond to the following questions by indicating how true or false each statement is about you. There are no correct or incorrect responses. Read each item carefully. Please think about how you feel about each item. Circle the letter that most closely corresponds to how the statements best describes your feelings. Use the following response scale to respond to each item. Your responses are confidential.	Use these codes: A-Always False B-Sometimes False C-Neutral D-Sometimes True E-Always True
1) I ask myself periodically if I am meeting my goals.	A B C D E
2) I consider several alternatives to a problem before I answer.	A B C D E
3) I try to use strategies that have worked in the past.	A B C D E
4) I pace myself while learning in order to have enough time.	A B C D E
5) I understand my intellectual strengths and weaknesses.	A B C D E
6) I think about what I really need to learn before I begin a task.	A B C D E
7) I know how well I did once I finish a test.	A B C D E
8) I set specific goals before I begin a task.	
9) I slow down when I encounter important information.	A B C D E
10) I know what kind of information is most important to learn.	A B C D E
11) I ask myself if I have considered all options when solving a problem.	A B C D E
12) I am good at organizing information.	A B C D E
13) I consciously focus my attention on important information.	A B C D E
14) I have a specific purpose for each strategy I use.	A B C D E
15) I learn best when I know something about the topic.	A B C D E
16) I know what the teacher expects me to learn.	A B C D E
17) I am good at remembering information.	A B C D E
18) I use different learning strategies depending on the situation.	A B C D E
19) I ask myself if there was an easier way to do things after I finish a task.	A B C D E
20) I have control over how well I learn.	A B C D E
21) I periodically review to help me understand important relationships.	A B C D E
22) I ask myself questions about the material before I begin.	A B C D E
23) I think of several ways to solve a problem and choose the best one.	A B C D E
24) I summarize what I've learned after I finish.	A B C D E
25) I ask others for help when I don't understand something.	A B C D E
26) I can motivate myself to learn when I need to.	A B C D E
27) I am aware of what strategies I use when I study.	A B C D E
28) I find myself analyzing the usefulness of strategies while I study.	A B C D E
29) I use my intellectual strengths to compensate for my weaknesses.	A B C D E
30) I focus on the meaning and significance of new information.	A B C D E
31) I create my own examples to make information more meaningful.	A B C D E
32) I am a good judge of how well I understand something.	A B C D E
33) I find myself using helpful learning strategies automatically.	A B C D E
34) I find myself pausing regularly to check my comprehension.	A B C D E
35) I know when each strategy I use will be most effective.	A B C D E
36) I ask myself how well I accomplished my goals once I'm finished.	A B C D E
37) I draw pictures or diagrams to help me understand while learning.	A B C D E
38) I ask myself if I have considered all options after I solve a problem.	A B C D E
39) I try to translate new information into my own words.	A B C D E
40) I change strategies when I fail to understand.	A B C D E
41) I use the organizational structure of the text to help me learn.	A B C D E
42) I read instructions carefully before I begin a task.	A B C D E

We would like you to respond to the following questions by indicating how true or false each statement is about you. There are no correct or incorrect responses. Read each item carefully. Please think about how you feel about each item. Circle the letter that most closely corresponds to how the statements best describes your feelings. Use the following response scale to respond to each item. Your responses are confidential.	Use these codes: A-Always False B-Sometimes False C-Neutral D-Sometimes True E-Always True
43) I ask myself if what I'm reading is related to what I already know.	A B C D E
44) I re-evaluate my assumptions when I get confused.	A B C D E
45) I organize my time to best accomplish my goals.	A B C D E
46) I learn more when I am interested in the topic.	A B C D E
47) I try to break studying down into smaller steps.	A B C D E
48) I focus on overall meaning rather than specifics.	A B C D E
49) I ask myself questions about how well I am doing while I am learning something new.	A B C D E
50) I ask myself if I learned as much as I could have once I finish a task.	A B C D E
51) I stop and go back over new information that is not clear.	A B C D E
52) I stop and reread when I get confused.	A B C D E

CURRICULUM VITAE

Christopher R. Rakes

Curriculum and Instruction
College of Education and Human Development
University of Louisville
Phone: (502) 852-4707
Fax: (502) 852-0139
Email: c.rakes@insightbb.com
Website: http://csrakes.yolasite.com

ACADEMIC BACKGROUND

2006 – Present	Ph.D., University of Louisville Major: Curriculum & Instruction – Secondary Mathematics Education Dissertation: *Misconceptions in Rational Numbers, Probability, Algebra, and Geometry* Advisor: Dr. Robert N. Ronau Expected Graduation Date: May 2010
2000	M.A., University of Kentucky Major: Secondary Mathematics Education
1999	B.A., University of Kentucky Major: Mathematics Education Minor: Biology

PROFESSIONAL EXPERIENCE

2009 – 2010	Graduate Teaching and Research Assistant University of Louisville
2008 – 2009	Doctoral Student Fellow University of Louisville
2007 – Present	Adjunct Faculty Kentucky Community & Technical College System
2004 – 2008	Secondary Mathematics Teacher Henry County High School

2001 – 2003 Adjunct Faculty
Kentucky Community & Technical College System

2000 – 2004 Secondary Mathematics Teacher
Eastside Technical Center

PUBLICATIONS

McGatha, M. B., Bush, W. S., & **Rakes, C. R.** (2009). The effects of professional development in formative assessment on mathematics teaching performance and student achievement. *Journal of Multidisciplinary Evaluation*, 6, 32-43.

In Press

Ronau, R. N., Rakes, C. R., Niess, M. L., Wagener, L., Pugalee, D., Browning, C., Driskell, S. O., & Mathews, S. M. (in press). New directions in the research of technology-enhanced education. In J. Yamamoto, C. Penny, J. Leight, & S. Winterton (Eds.), *Technology leadership in teacher education: Integrated solutions and experiences*. Hershey, PA: IGI Global.

Under Review

Rakes, C. R., Valentine, J., & McGatha, M. C. (2010). *Methods of Instructional Improvement in Algebra: A Systematic Review and Meta-Analysis.* Manuscript submitted for publication.

In Progress

McGatha, M. C., & **Rakes, C. R.** (2010). *Instructional coaching.* Manuscript in preparation.

Rakes, C. R., Ronau, R. N., Niess, M. L., Wagener, L., Pugalee, D., Driskell, S. O., & Harrington, R. (2010). *Research design in technology applications for mathematics education*. Manuscript in preparation.

Ronau, R. N., & **Rakes, C. R.** (2010). *Using the CFTK framework to examine mathematics teaching.* Manuscript in preparation.

Ronau, R. N., **Rakes, C. R.**, & Niess, M. L. (Eds.). (2011). *Educational Technology, Teacher Knowledge, and Classroom Impact: A Research Handbook on Frameworks and Approaches.* Hershey, PA: IGI Global.

Ronau, R. N., & **Rakes, C. R.** (2010). *A comprehensive framework for teacher knowledge: A lens for examining research.* Manuscript in preparation.

Unpublished Technical Reports

Rakes, C. R., & Rudasill, K. M. (2008, August). *Analysis of a district school satisfaction survey: Technical report* (Report No. 2). Louisville, KY: University of Louisville.

Rudasill, K. M., & **Rakes, C. R.** (2008, August). *Analysis of a district school satisfaction survey: General report* (Report No. 1). Louisville, KY: University of Louisville.

CONFERENCE PRESENTATIONS

National

Rakes, C. R. (2010, April). *Addressing mathematical misconceptions: Is probability the independent variable?* Paper presented at the annual meeting of the American Educational Research Association, Denver, CO.

Rakes, C. R., Ronau, R. N., Niess, M. L., & Wagner, L. (2010, April). *Research in mathematics instructional technology: Current trends and future demands.* Paper presented at the annual meeting of the American Educational Research Association, Denver, CO.

Rakes, C. R., & Ronau, R. N. (2010, April). *Research in mathematics instructional technology: Current trends and future demands.* Poster presented at the research presession of the annual meeting of the National Council of Teachers of Mathematics, San Diego, CA.

Rakes, C. R., Wagener, L., & Ronau, R. N. (2010, January). *New directions in the research of technology-enhanced education.* Paper presented at the annual meeting of the American Mathematics Teacher Educators, Irvine, CA.

Rakes, C. R. (2009, April). *Effective strategies for teaching algebra: A meta-analysis using hierarchical linear modeling.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Ronau, R. N., Wagener, L., & **Rakes, C. R.** (2009, April). A comprehensive framework for teacher knowledge: A lens for examining research. In R. N. Ronau (Chair), *Knowledge for Teaching Mathematics, A Structured Inquiry.* Symposium conducted at the annual meeting of the American Educational Research Association, San Diego, CA.

Ronau, R. N., **Rakes, C. R.**, Wagener, L., & Dougherty, B. (2009, February). *A comprehensive framework for teacher knowledge: Reaching the goals of mathematics teacher preparation.* Paper presented at the annual meeting of the Association of Mathematics Teacher Educators, Orlando, FL.

McGatha, M. B., Bush, W. S., & **Rakes, C. R.** (2009, February). *Formative assessment for middle-school mathematics teachers*. Paper presented at the annual meeting of the American Mathematics Teacher Educators, Orlando, FL.

Regional

Rakes, C. R. (2009, November). *Using virtual manipulatives to enhance students' understanding*. Paper presented at the regional meeting of the National Council of Teachers of Mathematics, Nashville, TN.

State/Local

Rakes, C. R. (2009, April). *Effective strategies for teaching algebra: A meta-analysis using hierarchical linear modeling.* Paper presented at the annual meeting of the Tri-University Spring Research Conference, Louisville, KY.

Rakes, C. R., & Brown, E. T. (2009, January). Virtual manipulatives in mathematics. Session presented at the quarterly meeting of the Greater Louisville Council of Teachers of Mathematics, Louisville, KY.

Rakes, C. R., & Walters, M. (2003, July). *Mathematics in Career and Technical Education*. Paper presented at the Kentucky Department of Education Career and Technical Education Summer Conference, Louisville, KY.

GRANTS

Research Teams

Geometry Assessments for Secondary Teachers (GAST; 2008 – Present). PI: Dr. William S. Bush. Co-PI: Dr. Robert Ronau. An NSF-funded project; $3,000,000 for 3 years. In this project, my role includes examination of geometry textbooks, assisting with in-class videotaping for pilot study, development of assessment instrument blueprint, and writing items for the assessment instrument.

Teacher Recruitment Effectiveness (2009 – Present). PI: Dr. Jeffrey C. Valentine. A project funded by the U. S. Department of Education ($90,000). In this project, our team is examining the characteristics of teachers collected at the time of hire in an urban, metropolitan school district and determining which characteristics are predictive of future success as measured by student achievement scores on the state-mandated assessment.

Internal Funds

University of Louisville Graduate Research Association Travel Grant, November 2009, awarded $420.

University of Louisville Graduate Research Association Travel Grant, April 2009, awarded $300.

University of Louisville Graduate Research Association Travel Grant, February 2009, awarded $300.

COLLEGE TEACHING EXPERIENCE

Mathematics Courses

College Algebra, MT 150 (Fall 2009; Spring 2009; Spring 2010). Jefferson Community & Technical College.

Elementary Algebra, MT 085 (Spring 2002). Maysville Community College/Central Kentucky Technical College.

Intermediate Algebra, MT 075 (2002, Spring). Maysville Community College/Central Kentucky Technical College.

Methods Courses

Methods for Teaching Mathematics, P – 5, EDTP 313 (2008, Fall). University of Louisville. Co-taught with Dr. E. Todd Brown.

Methods for Teaching Mathematics, P – 5, EDTP 604 (2008, Fall). University of Louisville. Co-taught with Dr. E. Todd Brown.

Teaching Mathematics with Technology, EDAP 397 (Fall 2009; Fall 2008). University of Louisville. Co-taught with Dr. Charles Thompson.

Research Courses

Evaluation and Measurement, ECPY 540 (Fall 2009). University of Louisville. Co-taught with Dr. Jill Adelson.

Hierarchical Linear Modeling, EDAP/ECPY 694 (Spring 2009; Spring 2010). University of Louisville. Co-taught with Dr. Robert Ronau and Dr. Thomas Tretter (Spring 2009) and with Dr. Jill Adelson (Spring 2010).

Introduction to Statistics, ST 291 (Fall 2009; Spring 2009; Spring 2010). Jefferson Community & Technical College.

Structural Equation Modeling, EDAP/ECPY 694 (Fall 2009; Fall 2008). University of Louisville. Co-taught with Dr. Robert Ronau and Dr. Thomas Tretter.

WORKSHOPS AND INSERVICE TEACHER TRAINING

Rakes, C. R. (2010, March). *Integrating technology into mathematics education to enhance equity.* College of Teaching and Learning, University of Louisville, KY.

Rakes, C. R. (2009, August). *Implementing a probability unit focused on developing conceptual understanding.* Bryan Station High School, Fayette County Public Schools, Lexington, KY.

Rakes, C. R. (2009, November). *Using vocabulary keyword strategies to enhance conceptual understanding.* Bryan Station High School, Fayette County Public Schools, Lexington, KY.

Rakes, C. R. (2009, November). *Targeting multiple learning styles through task rotation.* Bryan Station High School, Fayette County Public Schools, Lexington, KY.

CONSULTING ACTIVITIES

Bryan Station High School. (2009). *Consultant for development of mathematics department professional development program.* Lexington, KY: Fayette County Public Schools.

Henry County High School. (2008 – 2009). *Consultant for creation of database for student tardy management.* New Castle, KY: Henry County Public Schools.

Henry County High School. (2006 – 2008). *Consultant for unit planning database creation and management.* New Castle, KY: Henry County Public Schools.

Henry County High School. (2005 – 2008). *Consultant for management of multimedia checkout system*. New Castle, KY: Henry County Public Schools.

Henry County High School. (2005 – 2008). *Kentucky Teacher Internship Program resource teacher*. New Castle, KY: Henry County Public Schools.

Eastside Technical Center. (2003 – 2004). *Kentucky Teacher Internship Program resource teacher*. Lexington, KY: Fayette County Public Schools.

Eastside Technical Center. (2002 – 2004). *Consultant on Kentucky Department of Education Career and Technical Education Program Assessments.* Lexington, KY: Fayette County Public Schools.

Eastside Technical Center. (2000 – 2004). *Consultant for Microsoft Access database creation and management*. Lexington, KY: Fayette County Public Schools.

PROFESSIONAL ORGANIZATIONS

American Educational Research Association (Member, 2008 – Present)
- Reviewer for National Conference Submissions (2009, 2010)
- Graduate Student Liaison for University of Louisville (2008 – Present)
- Student Reviewer for *Journal of Teacher Education* under Dr. Robert N. Ronau (2009)

Association for Advancement of Computing in Education (Member, 2008 – Present)
- Reviewer for *Contemporary Issues in Technology and Teacher Education* (CITE) Journal, 2009 - Present
- Reviewer for Society for Technology and Teacher Education (SITE) annual conference (2010)
- Student Reviewer for Annual SITE conference under Dr. Robert N. Ronau (2007)
- Student Reviewer for *Contemporary Issues in Technology and Teacher Education (CITE)* Journal under Dr. Robert N. Ronau (2007)

Association of Mathematics Teacher Educators (Member, 2008 – Present)
- Reviewer for National Conference (2010)

National Council of Teachers of Mathematics (Member, 2000 – Present)
- Reviewer for *Journal of Research of Mathematics Education* (2009 – Present)
- Reviewer for *Mathematics Teacher* (2008 – Present)

Psychology of Mathematics Education, International and North American Chapter (Member, 2009 – Present)

United Nations Association of the United States of America (Member, 2006 – Present)

SERVICE

University of Louisville (2008 – 2010)
- College of Education and Human Development Technology Committee (Graduate Student Member)

- Department of Educational & Counseling Psychology Search Committee, Graduate Student Member (Psychometrician Position)

- University Graduate Student Council
    - Teaching & Learning Representative
    - Grievance Sub-Committee Member

- Tri-University Spring Research Conference Program Chair. (2009, Spring). Regional Research Conference sponsored by the University of Louisville, University of Kentucky, and University of Cincinnati.

- Tri-University Spring Research Conference Technology Committee Member. (2009, Spring). Regional Research Conference sponsored by the University of Louisville, University of Kentucky, and University of Cincinnati.

HONORS AND AWARDS

Graduate Dean's Citation Award, May 2010

University of Louisville Graduate Assistantship, 2009-2010

University of Louisville Graduate School Fellowship, 2008 – 2009

Golden Key International Honour Society, 2006 – Present

Eastside Technical Center: Outstanding Service Award, 2004

Kentucky SkillsUSA VICA: Outstanding Advisor of the Year, 2003

Who's Who among America's High School Teachers: Henry County High School, 2005 and 2007; Eastside Technical Center, 2002