Measuring Teacher Beliefs about Mathematics Discourse:

An Item Response Theory Approach

By

Heeju Jang

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Education

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Xiaoxia Newton, Chair
Professor Mark Wilson
Professor Maryl Gearhart
Professor Deborah Nolan

Fall 2010

Measuring Teacher Beliefs about Mathematics Discourse:

An Item Response Theory Approach

by Heeju Jang

Abstract

Measuring Teacher Beliefs about Mathematics Discourse:

An Item Response Theory Approach

By

Heeju Jang

Doctor of Philosophy in Education

University of California, Berkeley

Professor Xiaoxia Newton, Chair

The purpose of this dissertation is to develop an instrument that can measure teacher beliefs about mathematical discourse as a continuum by exploring and describing qualitatively distinct levels of teacher beliefs, namely, Univocal, Partial Univocal, Emerging Dialogical, and Dialogical. Prior research indicates the importance of understanding teachers' beliefs in the development of their teaching practice and the impact of their teaching practice in K-12 classrooms. However, assessing teacher beliefs has been difficult and often unsuccessful largely due to poor conceptualizations and measurement challenges associated with assessing beliefs. The field of teacher education is in need of carefully conceptualized and operationalized measures of teacher beliefs that are valid and reliable to understand relationships among teacher beliefs, teaching practice and student outcomes in order to improve instructional practice and inform educational policy.

This study reports the development of such a measure using the four building blocks recommended by Wilson (2005). The item design consisted of hypothetical teaching situations that present students' correct and incorrect thinking, and to then ask the teacher to respond in order to lead a mathematical discussion.

The participants in the study include a total of 168 pre-service teachers in 10 teacher education programs in the California State University system, the University of California system, and one private university, and 27 in-service teachers across the states. Results showed generally positive evidence for the validity and reliability of the measure. In addition, the findings of this dissertation suggest that the teacher belief about mathematical discourse (BMD) measure is a potentially promising tool for informing and designing elementary mathematics method courses. Suggestions for further research on the validity and reliability evidence are outlined

Table of Contents

Dedication

For mom, dad, and my loving husband, Foley.

CHAPTER 1: INTRODUCTION

The United States ranks far below other developed nations in K-12 mathematics achievement according to many international assessments such as the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA) (Gonzales, 2004). The poor achievement of U.S. students in mathematics is a complex problem that requires complex solutions. Some of the common factors for poor performance that emerge across studies are (a) teachers' weak preparation in mathematics (Goe & Stickler, 2008), teachers' lack of knowledge and skills about instructional practices (Darling-Hammond, 2000; Wenglinsky, 2000), and (c) teachers' beliefs about mathematics and pedagogy (Ambrose, 2004; Monk, 1994). This study addresses the last factor. Specifically, the study focuses on developing a new instrument to measure elementary teachers' beliefs about mathematical discourse.

Why Teacher Beliefs Matter

I begin my study with an examination of the research on teachers' beliefs and the research on mathematical discourse so as to understand and support the development of the instrument. Prior research indicates the importance of understanding teachers' beliefs, which play an important role in their acquisition of subject matter knowledge and the development of their teaching practice. Beliefs filter the teaching experiences of educators and influence their instructional practice (Hofer, 2001; Muis, Bendixen, & Haerle, 2006; Pajares, 1992). Seymour Sarason (1971) stated that educational change depends on what teachers do and think. Lortie (1975) also said one's personal predispositions stand at the core of becoming a teacher. While curriculum materials and instructional resources might suggest new directions in mathematics education, implementing any reform depends on what teachers believe about a subject matter, and teaching and learning in general. Early beliefs about teaching mathematics, according to past research, are very resistant to change (Cooney, Shealy, & Arvold, 1998; Cross, 2009), and remain virtually unchanged over time despite later experience and education training (Pajares, 1992). To make the educational training of preservice teachers more effective, diagnosing their beliefs is important to affecting them. Understanding preservice teachers' beliefs should be an essential part of designing the curriculum and pedagogy in teacher education programs to effectively address the challenges associated with changing ingrained beliefs.

Although understanding teachers' beliefs is important for understanding their instructional practice, the research in this area is far from presenting a clear picture. Belief is a messy construct, and is often used interchangeably with terms such as "attitude," "value," "opinion," "perception," "conception," "preconception" and "disposition." In this study, the definition of belief draws from definitions by two scholars, Rokeach and Dewey. Rokeach (1968) defined belief as "any simple proposition, conscious or unconscious, inferred from what a person says or does" (p.113). Dewey (1933) believed that the definition "covers all the matters of which we have no sure knowledge and yet which we are sufficiently confident of to act upon and also the matters that we now accept as certainly true, as knowledge, but which nevertheless may be questioned in the future…." (p.6). Drawing from these definitions by Rokeach and Dewey, this study proposes the following operational definition of teacher beliefs about mathematical discourse: *The conscious and unconscious ideas and thoughts teachers have about how teachers and students should participate in classroom discussion to build mathematical knowledge.* Belief is a latent construct that cannot be directly observed, which makes its measurement challenging.

One way to understand a teacher's belief, used in this study, is to provide him/her with hypothetical teaching situations that present students' correct and incorrect thinking, and to then ask the teacher how he or she would respond in order to lead a mathematical discussion. From a teacher's response, his/her belief can be inferred.

Mathematical Discourse

The purpose of this dissertation is to develop an instrument to measure teachers' beliefs about mathematical discourse. In this section, to clarify belief, the construct of interest, I present different types of mathematical discourse conceptualized by Brendefur and Frykholm (2000), and Mortimer and Scott (2003). These two sources are valuable resources for the belief construct that my instrument is designed to measure. Each framework uses different labels to describe different types of mathematical discourse, yet mathematical discourse is conceptualized as a continuum. The framework of Brendefur and Frykholm (2000) and the framework of Mortimer and Scott (2003) both move beyond dichotomized thinking and explore a range of perspectives on how teachers interact with students in the process of developing ideas in the classroom.

Measuring Teacher Beliefs about Mathematical Discourse

The links among teacher preparation context for learning, what preservice teachers learn, how their beliefs change, and how teacher learning is played out in practice in K-12 classrooms are crucial to investigate. The effect of teaching practice on students' learning has been an important research question in the field of research on teacher education (Cochran-Smith & Zeichner, 2005). Yet, we know little about this relationship. The investigation of teachers' beliefs often uses case studies with small sample sizes, and this kind of study tends to be interpretive in nature, aiming for particularity instead of generalization (Erickson, 1986). The lack of a shared conceptual framework in interpretive research makes it challenging to aggregate findings across studies, and creates difficulties when attempting to compare results across studies in order to improve instructional practice and inform educational policy (Borko, Liston, & Whitcomb, 2007). In an effort to address the limitations of qualitative research, some studies use surveys with Likert-scale items. Surveys that employ Likert-scale items provide a consistent measurement tool and one way to gather information from many participants quickly and easily. However, surveys that employ these items also have their own limitations associated with the interpretation of respondents' answers. Thus the field is in need of valid and reliable measures to investigate these important relationships (Cochran-Smith & Zeichner, 2005). A carefully conceptualized and operationalized measure of teacher belief can play an important role in larger studies of relationships among teacher beliefs, classroom practice, and student outcomes (Cochran-Smith & Zeichner, 2005).

The debate around mathematical instruction makes the development of a valid and reliable measure challenging because the beliefs of teachers have been polarized in a dichotomous, simplistic way as either reform-based or traditional (Boaler, 2008; Schoenfeld, 2004). This dichotomy is problematic because neither camp has clear-cut identifying characteristics (Boaler, 2008). The experts in the field have, in recent years, agreed that mathematics instruction should move toward a more centrist position that utilizes both approaches (Benbow & Faulkner, 2008; Boaler, 2008; Lobato, Clarke & Ellis, 2005; Sherin, 2002). This study will contribute to the field of teacher education by developing a valid and reliable measure of teachers' beliefs about mathematical discourse that goes beyond the dichotomous way of categorizing teachers' beliefs, and instead describes teachers' beliefs as a continuum.

The measure developed in this study focuses on teachers' beliefs about mathematical discourse. Communication has been recognized as essential to reform-oriented teaching of mathematics by the National Council of Teachers of Mathematics (NCTM, 1991; 2000; 2006). According to the Principles and Standards for School Mathematics (PSSM), students should be able to communicate their mathematical thinking coherently, precisely, and clearly to peers, teachers, and others using mathematical language, and be able to analyze and evaluate the mathematical thinking and strategies of others. The vision of the standards reflects the change in the role of mathematics teachers. Unlike the traditional role of the teacher who is a "dispenser of knowledge" (Stein, Engle, Hughes, & Smith, 2008), the new role envisioned for mathematics teachers is to orchestrate mathematical discussions, with the goal of understanding and extending students' thinking (Hufferd-Ackles, Fuson, & Sherin, 2004). By having a measure to identify where a teacher's belief about how to lead a mathematical discussion falls on a continuum, educational researchers can be one step closer to investigating the relationship among teachers' beliefs about mathematical discourse, instructional practice, and students' learning in K-12 classrooms.

In order to develop a valid and reliable measure, I gathered and analyzed different types of evidence for validity and reliability. Validity refers to the degree to which an instrument accurately reflects or assesses the specific concept that the researcher is attempting to measure (AERA, APA, & NCME, 1999). This study addresses questions about the validity of the instrument through construct, content response processes, internal structure, and relations to other existing measures of identical or similar constructs (Wilson, 2005). Reliability is concerned with the accuracy of the actual measuring instrument or procedure (AERA, APA, & NCME, 1999). This study addresses questions about the reliability of the instrument through the use of Cronbach's alpha and the Pearson separation reliability (Cronbach, 1990).

Significance of this Research

One of the important contributions of this study is its utility as a measure in the field of teacher education and professional development. It will be informative for teacher educators to understand what preservice teachers believe when they enter their program and how preservice teachers' beliefs change as they progress through the program, complete the program, and enter the teaching force. It will also be informative for teacher educators to understand how preservice teachers' beliefs affect their teaching practice, or vice versa, and systematically investigate whether, and in what specific ways, their program can make an impact on preservice teachers' beliefs and teaching practices. If there are any inconsistencies between teachers' beliefs and their instructional practice, sources of inconsistency should be identified by teacher educators to better understand the relationship between beliefs and practice to support teachers' learning to teach.

This study will also demonstrate the development of a valid and reliable measure using the framework of Item Response Theory (IRT), which will be a novel contribution to the field of research on teacher education. IRT is the study of test and item scores based on assumptions concerning the mathematical relationship between hypothesized traits and item responses. Modeling the relationships between ability and a set of items has advantages over methods used by classical measurement theorists. Since IRT models have not been utilized in research on teacher beliefs so far, this study will introduce a measure that is psychometrically more rigorous than other existing measures with similar goals in the field. A brief introduction of IRT will be outlined in Chapter 3. Overall, by providing a way to explore variations in teacher beliefs, this study will not only produce a useful tool based on existing research but also present a more developed and refined language to talk about teacher beliefs to further advance the discussion about teacher belief change and its connection with instructional practices.

Research Questions.

The purpose of this study is to develop an instrument that can measure teacher beliefs about mathematical discourse, and establish evidence for validity and reliability of the instrument. The research study will answer the following questions:

1. What is the evidence for content validity?
2. What is the evidence for response process validity?
3. What is the evidence for internal structure validity?
4. What is the evidence for external validity?
5. What is the evidence for reliability?
6. To what degree, if any, are teacher beliefs about math discourse associated with other factors such as demographics, educational background and teaching experience?

CHAPTER 2: REVIEW OF LITERATURE

Communication in Mathematics Education

Within the mathematics education community there is strong interest in the use of discussion and conversation in teaching and learning mathematics (Atkins, 1999; National Council of Teachers of Mathematics, 1991, 2000; OECD Program for International Student Assessment, 2003; Schifter, 1996). According to the Communication Principle in the Principles and Standards for School Mathematics (PSSM) (NCTM, 2000), instructional programs from prekindergarten through grade 12 should enable students to organize and consolidate their mathematical thinking through communication in a coherent, precise, and clear way. Instructional programs should also enable students to analyze and evaluate their own mathematical strategies and those of their peers using mathematical language. In short, they should be able to engage in mathematical discourse.

Discourse in this case means either "written or spoken communication or debate" or "a formal discussion.” Discourse makes thinking public and creates an opportunity for the negotiation of meaning and agreement (Bauersfeld, 1995). At the same time, discourse provides collective support for developing one's thinking, drawing it out through the interest, questions, probing, and ideas of others (Cobb, 1992; 1997; Krummheuer, 1995; Wood, Cobb, & Yackel, 1991). In this study, I define discourse as the way that teachers interact with students in the process of developing mathematical ideas during open discussion.

Instructional programs should be designed with the goal of guiding students’ engagement in mathematical discourse, and the goal can be achieved by teachers who organize and shape the classroom. With regard to mathematical communication, it is the role of teachers to create a safe classroom culture in which students can communicate with one another as they work through mathematical problems (Stein, Engle, Hughes, & Smith, 2008; Stigler, & Hiebert, 1997). In this setting of small groups and whole-class discussions, students can share their strategies and solutions, using explanations, justifications, and arguments (Lappan, 1997; Stein, Engle, Hughes, & Smith, 2008). Teachers should be able to listen and respond to student thinking while encouraging other students to participate in the ongoing mathematical discourse. This kind of teaching requires teacher flexibility, because different responses are necessary to meet student needs, depending on what students say or do, and what they understand. The goal is no longer simply getting the correct answers to mathematics problems, but helping students to develop, clarify, extend, and communicate their mathematical thinking (Hufferd-Ackles, Fuson, & Sherin, 2004).

Student learning and motivation benefit a great deal from participation in certain types of mathematical discourse (Ball, 1993; Hiebert, 1996; Hufferd-Ackles, Fuson, & Sherin, 2004; Lampert, 1990; NCTM, 1991, 2000; Silver & Smith, 1996). Mathematics can be viewed as a language requiring specific literacy practices. In this view, mathematical discourse is a way to understand and express mathematical ideas (NCTM, 2000). When students clarify their mathematical thinking orally or in writing, students learn to think clearly for themselves and present their thinking clearly to others much as theoretical and applied mathematicians do in their practice. When students listen to, share, and discuss their mathematical thinking with one another it gives them the opportunity to learn from one another and to learn to communicate with one another mathematically (Ball, 2005; NCTM, 1989; 2000). Communicating with one another mathematically means that students use the signs, symbols, and terms of mathematics in a problem solving situation as they communicate their ideas. In the process, students learn to clarify, refine, and consolidate their mathematical thinking (NCTM, 1989). Thus current recommendations by prominent mathematics educators place considerable emphasis on the role of classroom discourse in supporting students' conceptual development and reasoning (Chapin, O'Connor, & Anderson, 2003; Cobb, Boufi, McClain, & Whitenack, 1997; Hiebert et al., 1996).

Mathematics Discourse in the Classroom: Tensions and Challenges

Traditional vs. 'reform' discourse.

The dominant form of discourse in the mathematics classroom has traditionally been unidirectional, or univocal, communication (Peressini & Knuth, 1998; Thompson, 1992). In this style, teachers convey information to students mainly by explaining mathematical concepts and modeling computational procedures that arrive at the correct answers. Teachers who use unidirectional communication tend to immediately correct students' wrong answers and explain why the answers are incorrect. Unidirectional communication is common in traditional mathematics classrooms in which teachers are viewed as the sole authorities of mathematics knowledge (Rittenhouse, 1998). In such a classroom, teachers typically play the role of "dispenser of knowledge" (Stein, Engle, Hughes, & Smith, 2008), transmitting knowledge to, and validating answers for, students who are expected to learn alone and in silence (Chazan & Ball, 1999; Silver & Smith, 1999).

In the 1990s, reformers of mathematics education called for a new classroom discourse in which students are offered opportunities to solve challenging problems and present and discuss their interpretations and solutions (Brendefur & Frykholm, 2000; Cobb et al., 1997). In this view, teachers are tasked with encouraging students to contribute to the learning process more actively by justifying mathematical thinking to one another, and questioning one another in a pattern of whole class discourse that is facilitated but not directed by teachers. The mathematics under consideration can even be altered through this process (Lampert, 1990; Steffe & D'Ambrosio, 1995). Thus the vision of reform-oriented mathematics discourse departs significantly from traditional classroom discourse (Cobb, Wood, Yackel, & McNeal, 1992), and is much more challenging to facilitate. In order for students to benefit from such a mathematics discussion, teachers need to rethink the forms and functions of discourse, and students need to learn new roles and new purposes for discussion in the context of learning mathematics.

Challenges to implementation of reform discourse.

Many teachers learned mathematics in traditional classrooms in K-12. Little discussion took place in these settings, and the teachers thus lack personal experience with discussions in math classrooms. Regardless of their experiences as elementary students, preservice teachers are also learning to teach for the first time (Ebby, 2000; Frykholm, 1996, 1999; Raymond, 1997), and mathematical discourse is a very challenging practice, whether it is univocal or not. Their lack of experience and the challenging learning curve in their chosen profession are significant barriers to them fully embracing open mathematical discourse as part of their classroom organization.

Another barrier to the implementation of mathematical discussion is the weak preparation of teachers in the field of mathematics. According to Leitzel, "the mathematical preparation of elementary school teachers is perhaps the weakest link in our nation's entire system of mathematics education" (Hungerford, 1994). In most states, teachers in grades K-6 are not mathematics specialists. Only 7 percent of elementary school teachers majored or minored in mathematics or mathematics education (National Science Board, 2004). Moreover, 40 percent of elementary school teachers report that they do not feel qualified to teach the mathematics content that they are charged with (Office of Postsecondary Education, 2004). Given the weak preparation of teachers in mathematics and their limited experiences with reform practices, teaching a reform-based mathematics curriculum is extremely challenging (Ball, 1996; Battista, 1993).

The barrier that is the focus of this dissertation, however, is within teachers themselves. Teachers' own deeply rooted beliefs about mathematics and mathematics teaching impede effective mathematical discussion. Research shows that teachers' past experiences with mathematics learning during K-12 shape their beliefs developed over years of schooling and educational experiences (Belenky et al., 1986; Cross, 2009; Handal, 2003). Beliefs about mathematics and teaching of mathematics are often formed and solidified well before prospective teachers enter preparation programs, and beliefs tend to remain intact despite their teacher education training (Torff & Warburton, 2005).

Many teachers are reluctant to accept discussion as a method of teaching mathematics because they tend to believe that the learning of mathematics takes place when teachers disseminate knowledge, facts, and algorithms, and they generally expect students to replicate the fields of knowledge disseminated (Boaler, 2002; Brooks & Brooks, 1993). Teachers may rely heavily on textbooks (Ben-Peretz, 1990) and claim that students should be completing the tasks in that medium, not inquiring about the mathematics in discussions (Morse, 1998). If these teachers observe students talking with one another or asking questions, they infer that the basics and other important mathematics are not being taught (Stiff, 2000). They believe independent practice using workbooks, rather than cooperation with groups, helps students learn the accepted concepts. These teachers seek one "correct" answer from students, and they provide the explanation as to why the answer is correct or incorrect. Thus the exploration or construction of new knowledge, for teachers who hold these beliefs, is not as highly valued as the ability to demonstrate mastery of conventionally accepted understandings (Brooks & Brooks, 1993).

Given these deeply rooted beliefs about mathematics and pedagogy, it is a challenge to encourage teachers to implement classroom discussion as a way to teach mathematics (Sherin, 2002). Beliefs about mathematics and pedagogy are hard to change even with formal preservice training or interventions through professional development. Due to the central role of mathematical discussion in reform visions, it is important to consider preservice teachers' beliefs about mathematical communication when designing such programs (Brendefur & Frykholm, 2000) with the goal of helping teachers to develop practices that are more consistent with reform visions (Blanton, 2002; Wilcox, Schram, Lappan, & Lanier, 1991). In the next section, I will further discuss why teacher beliefs are important to consider in designing preservice and professional development programs.

The Role of Beliefs in the Teaching of Mathematics

Mathematics education researchers have become increasingly aware of the influence of teacher beliefs as more emphasis has been placed on the role of teaching in the learning process (Cobb et al., 1991; Nesper, 1987; Pajares, 1992; Philipp, 2007; Philipp et al., 2007; Raymond, 1997; Torff, 2005; Wilson & Cooney, 2002). Research into the beliefs of teachers is important to help them develop as self-regulated, critically reflective professionals (Ng, Nicholas, & Williams, 2010). Seymour Sarason (1971) suggested that educational change depends on what teachers do and think. While curriculum reform might suggest new directions, like more discussion, in mathematics education, the implementation of these new directions depends in part on what teachers believe about mathematics in general, teaching, and the learning of mathematics. Reforming the instructional practices of many mathematics teachers can be only actualized if we better understand teachers' beliefs, and how beliefs are related to practice (Cross, 2009). This dissertation focuses on teacher beliefs with the goal of understanding their qualitatively different variations to attend to the process of learning (discussion) as well as to the content to be learned (mathematics) (Klatter, Lodewijks, and Aarnoutse, 2001).

So what do we know about beliefs? Beliefs are studied in diverse fields, and, as a result, the body of work on beliefs has produced a variety of meanings for the term (Abelson, 1979; Dewey, 1933; Rokeach, 1968).There is no universal definition that is agreed upon (Cross, 2009). Beliefs are defined as “embodied conscious and unconscious ideas and thoughts about oneself, the world, and one’s position in it, which are considered by the individual to be true” (Pajares 1992; Thompson 1992). Rokeach (1968) defines beliefs as “any simple proposition, conscious or unconscious, inferred from what a person says or does.” Dewey (1933) argues that belief “covers all the matters of which we have no sure knowledge and yet which we are sufficiently confident of to act upon and also the matters that we now accept as certainly true, as knowledge, but which nevertheless may be questioned in the future… (p. 6).” Drawing from the definitions by Cross, Rokeach, and Dewey have constructed, an operational definition of teacher beliefs about mathematical discourse as follows: *The conscious and unconscious ideas and thoughts teachers have about how teachers and students should participate in classroom discussion to build mathematical knowledge.* I will argue beliefs as I define them can be inferred from what teachers do when leading a mathematical discussion in the classroom and from what they report when discussing their plans for discussion or describing their methods of facilitating discussion.

A substantial body of research on teacher beliefs suggests that teachers are crucial change agents in educational reforms, and that changing teacher beliefs is a precursor to changing their teaching practice (Ernest, 1989; Fang, 1996; Stipek, Givvin, Salmon, & MacGyvers, 2001; Thompson, 1992). Beliefs are personal and stable, and considered to be influential in determining how individuals frame problems and structure tasks (Rimm-Kaufman & Sawyer, 2004). Research has also consistently illustrated ways in which teachers translate their knowledge of mathematics and pedagogy into practice through the filter of their beliefs (Manouchehri, 1997; Thompson, 1992). Teachers’ instructional practice is often based on what they believe the subject matter is and how it should be taught (Boaler, 2008; Laurenson, 1995). Preservice teachers bring with them to teacher education courses a set of beliefs that are firm and resistant to change (Murphy, Delli, & Edwards, 2004). Preservice teachers’ beliefs are influenced by their experience as students and their beliefs surface when they start teaching (Ng, Nicholas, & Williams, 2010). The relationship between beliefs and practice is not one-directional (Gusky, 1986), and the connection between what teachers do and what teachers think is a complex relationship; what they do affects what they think and what they think affects what they do. Changing beliefs about instruction and classroom practice is a complex task (West & Staub, 2003). Thus teachers’ beliefs must be addressed in conjunction with engaging teachers with the implementation of new instructional practice, providing a context for reflecting on those practices and revising beliefs (Benbow, 1993; Britzman, 1991; Doerr & Lesh, 2003).

Measuring Teacher Beliefs

Given the importance of mathematical discussion, and the challenges associated with changing teacher beliefs about mathematical discussion, it is important to develop ways to measure teacher beliefs about mathematical discussion. The purpose of this dissertation is to develop an instrument to do just that. Such an instrument will be a contribution to the field of teacher education and teacher professional development because it will provide a way to establish baseline data on beliefs and thus provide a gateway to changing them. In this section, I will discuss various ways to measure teacher beliefs about mathematical discourse.

Currently, there are several ways to measure teacher beliefs, each with strengths and weaknesses. Surveying is a classic method of data collection. The advantage of using surveys is that they are easy to administer, and collection of large amounts of data in a timely fashion is possible. The Likert scale is the most frequently used survey item type. The most generic form of the Likert scale is the provision of a stimulus statement, often called a stem, and a set of standard options among which the respondent can choose (e.g., strongly agree, agree, disagree, and strongly disagree, with a middle neutral option). Likert scale items allow a respondent to provide slightly more feedback than is possible with a simple close-ended question (e.g., yes, no), and that feedback is much easier to quantify than an open-ended response.

Although a survey using a Likert scale is a common approach widely used to assess attitudes or beliefs, it has several limitations. First, there is little guidance provided to respondents in judging the difference between strongly disagree and disagree, or agree and strongly agree, and thus respondents may have different ideas about these distinctions (Wilson, 2005). Second, it is difficult to know how a respondent interprets survey items, because most survey items tend to be general statements that provide unified context. Therefore respondents generally have to rely on their own personal reference in order to address the survey items (Ambrose, Clement, Philipp, & Chauvot, 2004). This is a serious threat to the validity of any instrument. Consider for example, the survey item, “It is important for a child to be a good listener in order to learn how to do mathematics” (Fennema, Carpenter, & Loef, 1990). In order to respond to this survey item, some respondents may wonder whether the item is asking about the importance of listening or the relative importance of listening compared to verbally articulating their thinking with their peers. Depending on the intention of the survey item, a respondent may give an answer that doesn’t accurately address the question. It is possible that different respondents will answer the same survey item with different situations in mind. The goal of this study is to develop and validate a survey that is more valid and reliable by developing questions that better elicit meaningful and interpretable information about teachers’ beliefs about mathematical discourse.

Measuring Beliefs about Mathematical Discourse

To measure teachers’ beliefs about mathematical discourse, we need to develop questions about communication between teachers and students, as well as among students. Currently there are a limited number of survey instruments designed to measure teacher beliefs about mathematical discourse. In these surveys, the word ‘beliefs,’ ‘perceptions,’ ‘views’ and ‘conceptions’ are often used interchangeably. One survey instrument (DeFranco, Gorgievski, & Truxaw, 2008) was developed to measure K-8 mathematics teachers’ perceptions about discourse in mathematics classes. The 5-point Likert-type survey, with 18 items, seeks to measure teachers’ perceptions of their use of dialogic (to construct new meaning), univocal (conveying information), and general discourse in their mathematics classes. A similar instrument is the Preservice Teachers’ Attitude about Discourse in the Mathematics Classroom, or PADM (Casa, McGlvney-Burelle, & DeFranco, 2007). This 5-point Likert-type survey, with 25 items, is designed to measure preservice teachers’ attitudes regarding discourse in the K-12 mathematics classroom. The PADM instrument purports to measure three dimensions: promoting mathematical reasoning, examining complex mathematical concepts, and valuing students’ mathematical ideas.

These surveys are easy to use to assess teachers’ attitudes toward mathematical discourse. However, these surveys were developed based on the assumption that two dichotomous approaches to mathematical discourse exist -- the dialogical vs. the univocal approach. I argue that these two approaches to mathematics discourse are not necessarily mutually exclusive, and are far from having clear-cut identifying characteristics. Thus it is important to develop an instrument that can measure qualitative differences along the continuum from univocal to dialogical as an alternative way of understanding teacher beliefs about mathematical discourse.

Misleading definitions and misguided debates.

The dichotomous way of conceptualizing mathematical discourse (univocal vs. dialogic) is prevalent in the broader context of debates about mathematics education. In this section, I review current discussion around mathematical instruction to support my position that it is important to move beyond dichotomies when developing measures of teachers' beliefs about mathematics pedagogy.

The dichotomous way of conceptualizing mathematics pedagogy is reflected in different approaches to teaching mathematics. For example, "teacher-directed" instruction, a "behaviorist" approach, or a "traditional" approach are all terms used interchangeably to describe one mathematics pedagogy whereas the "student-centered" approach, "constructivist" approach, or "reform-based" approach are terms used to describe another mathematics pedagogy. However, the definitions of these terms are unclear, and clear conceptualizations are lacking. Each of these terms has multiple meanings, and the meanings are often not operationalized or supported with illustrative examples (Lobato, 2008).

Different definitions for the same terms have been employed by different groups of scholars. For example, the National Mathematics Advisory Panel (2008) defined the terms 'teacher-centered instruction' and 'student-centered instruction' to distinguish different approaches to teaching mathematics. The Panel (2008) used "teacher-centered instruction" to refer to occasions when "teachers facilitate, encourage, and coach but do not explicitly instruct by showing and explaining how things work," whereas "student-centered" instruction is defined as "the students doing the teaching of the mathematics" (Boaler, 2008). However, these terms are understood slightly differently by other experts in the field of mathematics education (Boaler, 2008). Teacher-directed instruction is generally understood as a teacher presenting methods to students who watch, listen, and then practice the methods (Boaler, 1998; Good & Grouws, 1997). On the other hand, student-centered instruction generally implies an approach in which learners are given opportunities to offer their own ideas and to become actively involved in their learning (Cobb, 1992; Confrey, 1990).

Student-centered instruction is often associated with terms and ideas like progressive instruction, a reform-based approach, the active learning of students, a constructivist approach, or inquiry-oriented instruction. On the other hand, teacher-centered instruction is associated with more traditional and behaviorist approaches largely focusing on skills and procedures, which tend to foster passive learning of students (Gales, & Yan, 2001). Student-centered instruction is often synonymous with the constructivist approach, and similarly, teacher-centered instruction with behaviorist approach. For example, some teachers might be described as traditional just because they lecture and have students work individually, whereas other teachers might be described as constructivist because they have their students work in groups (Boaler, 2008). However, neither a lecture nor group work is definitive of either approach.

Thus terms and constructs used to contrast approaches to mathematics teaching are not necessarily mutually exclusive, and they are far from having clear-cut characteristics despite common associations with superficial teaching features. I argue that all teaching methods, regardless of their position at one pole or another of any given dichotomy, may be effective in supporting student learning. It is possible that students are engaged in interesting mathematical inquiries during lecture, while a lack of appropriate scaffolding during small group activities may result in no fruitful discussion among students. Thus assumptions about teaching effectiveness based on superficial features of teaching do not result in meaningful analysis of mathematics teaching, or beliefs about mathematics teaching. Although some scholars argue that the field of research on mathematics instruction has moved beyond dichotomized thinking altogether (Lobato, Clarke, & Ellis, 2005; Sherin, 2002), this progress is not reflected yet in the current measurement practice. It therefore makes sense to develop instruments that that reflect this progress to measure teacher beliefs.

Here, I present these two frameworks as examples of conceptualizing the construct of interest as a continuum instead of as a dichotomy. The two frameworks are organized slightly differently, yet each contributes valuable perspectives on how teachers interact with students in the process of developing ideas in the classroom.

Brendefur and Frykholm.

Brendefur and Frykholm (2000) describe mathematical discourse as it develops from a simple form to more complex forms, and organizes the perspectives of mathematical communication into four categories: unidirectional, contributive, reflective, and instructive. The pattern of communication in the first two levels (unidirectional, and contributive) may be described as univocal, and the pattern of communication in the last two levels (reflective and instructive) may be described as dialogical (Peressini & Knuth, 1998; Wertsch & Toma, 1995). Univocal is represented by “the transmission model and presupposes that a single, univocal message is transmitted from sender to receiver” (Wertsch & Toma, 1995). Thus a conduit metaphor is helpful to imagine univocal discourse — knowledge is sent in one direction.

Contributive communication focuses on interactions among students, as well as interactions between teachers and students. Teachers may provide opportunities for students to discuss mathematical tasks with one another, present solution strategies, or assist each other in developing solutions and appropriate problem solving strategies. Yet the conversation is limited to teachers providing assistance to arrive at the correct answer, and conversations are typically corrective in nature.

Common in classrooms, unidirectional communication and contributive communication are both discussions in which teachers dominate by lecturing, asking closed questions, and allowing few opportunities for students to communicate their strategies, ideas, and thinking. In classrooms with unidirectional or contributive communication, mathematics is treated as a static body of knowledge, and the conversations are predominantly focused on teachers explaining with the goal of transmitting a body of knowledge. The important distinction between unidirectional and contributive communication depends on the social norms of the classroom and the expectations of how knowledge is conveyed (Cobb, Wood, Yackel, & McNeal, 1992; Kazemi, 1998; Kazemi & Stipek, 1997). In the former, teachers are assumed to be the authority of mathematical knowledge, whereas in the latter, teachers value students’ articulation of their mathematical understanding although they still retain mathematical authority.

Reflective communication is based on a more complex level of communication. Teachers at this level provide opportunities for students to reflect on the relationships among the mathematical topics by sharing ideas and strategies. Reflective communication is similar to contributive communication in that students share their ideas, strategies, and solutions with peers and teachers. However, teachers at this level elicit students’ mathematical ideas, and use them as springboards for deeper investigations and explorations in which “repeated shifts [occur] such that what the students and teachers do in action subsequently becomes an explicit object of discussion” (Cobb, Boufi, Mclain, & Whitenack, 1997). Thus the move from contributive to reflective communication is critical because the purpose of students sharing their ideas is not limited to verifying the correctness of their answers, but extends to deepening students’ mathematical understanding.

Finally, the last level of communication is instructive communication. At this level, teachers provide students the opportunity to make their thinking public so that teachers begin to understand "the thought processes, strengths, and limitations of particular students" and incorporate "students' mathematical ideas and conjectures into the instructional sequence" (Brendefur and Frykholm, 2000). As a result, students' understanding of the mathematics at hand is deepened, and "classroom communication becomes instructive" (Brendefur and Frykholm, 2000). Teachers at this level of communication allow the path of the classroom progression to be altered based on the conversations in the classrooms.

At the level of reflective and instructive communication, the focus of the conversation is to generate meaning (Peressini & Knuth, 1998; Wertsch & Toma, 1995). Students are asked to not only share their methods and answers but incorporate their peers' ideas into their own and build upon conversations in meaningful ways. Brendefur and Frykholm developed an important theoretical framework that looks at classroom discourse in continuum with four points of measurement. I take their work as one basis upon which to develop the construct of interest in this dissertation.

Scott and Mortimer.

The framework presented by Scott and Mortimer (2005) is another important contribution to the field, because it explores qualitatively different categories that describe how teachers interact with students in the process of developing ideas in the classroom. Their framework is organized based on multiple dimensions.

Scott and Mortimer (2005) characterized the ways in which teachers guide meaning-making interactions. Central to this framework is the concept of a communicative approach which consists of two dimensions: the dialogic-authoritative dimension, and the interactive-noninteractive dimension. The different classes of communicative approach are defined in terms of whether the classroom discourse is authoritative or dialogic in nature and whether it is interactive or noninteractive (Scott & Mortimer, 2003). Thus four classes of communicative approach are generated along each of two dimensions (Table 1).

	Interactive	*Noninteractive*
Dialogic	A. Interactive/Dialogic	B. Noninteractive/Dialogic
Authoritative	C. Interactive/Authoritative	D. Noninteractive/Authoritative

Table 1 *Four Classes of a Communicative Approach (adapted from Scott & Mortimer, 2005).*

An authoritative function of communication is equivalent to unidirectional communication in which the purpose of communication is to focus the students' full attention on just *one* meaning (e.g., the correct answer or the standard method). In contrast, dialogic according to Scott and Mortimer means recognizing students' points of view and taking into account a range of ideas. Dialogic interaction can be played out with different levels, however (Bakhtin, 1981). For example, a teacher might simply ask for the students' points of view and list them on the board. The discourse is open to different points of view, but in this case, the teacher may not attempt to utilize those views through comparing and contrasting. In another example, a teacher might elicit the students' points of view, list them on the board, encourage students to agree or disagree with particular points, and then take it one step further by facilitating the discussion to connect how the ideas relate to one another. Both approaches are dialogic in the sense that they encourage different ideas to be presented, but there is qualitative difference in what each teacher did with students' shared ideas to shape the discussion.

Authoritative discourse, however, neither encourages exploration of different points of view nor brings them together. Teachers may reshape or ignore students' ideas or questions when they do not exactly fit in the development of knowledge in the teachers' mind. Authoritative discourse is closed to the points of view of others. Acknowledgment and exploration of different perspectives is absent or limited. What makes a conversation functionally dialogic is that different ideas are acknowledged, rather than whether an idea is produced by a group of people or an individual. This point leads to the second dimension of the communicative approach: that the conversation can be interactive in the sense of allowing for the participation of more than one person or noninteractive in the sense of excluding the participation of other people. Combining the two dimensions, any episode of classroom dialogue can be identified as being either interactive or noninteractive on the one hand, and dialogic or authoritative on the other. This framework is organized in terms of two dimensions although each dimension accepts a dichotomy as its basis. This framework is important to form the basis of the measurement of teacher beliefs because it identified two distinct dimensions that can be further explored and developed on a continuum.

Integration and application.

The development of a measurement instrument in this study was informed by both the framework developed by Brendefur and Frykholm (2000) and the framework developed by Scott and Mortimer (2005). The framework by Scott and Mortimer (2005) was useful in identifying the two important dimensions that matter to mathematical discourse. However, it was less useful in that each dimension was treated in dichotomy instead of on a continuum. These dimensions of mathematical discourse are not necessarily mutually exclusive, and are far from having clear-cut identifying characteristics. Thus it is desirable to develop an instrument that can measure qualitative differences along the continuum from univocal to dialogical as an alternative way of understanding teacher beliefs about mathematical discourse. The framework developed by Brendefur and Frykholm (2000) was helpful because it was organized from simple to more complex forms, in a continuum, yet characteristics of each level were not as clear as they could be. Drawing from the strengths of both frameworks I developed the construct map of mathematical discourse presented in Chapter 3.

Having reviewed the literature on teacher beliefs and the existing measurement approach using surveys with Likert-scale items, I have supported the need for a more valid and reliable survey designed to measure teacher beliefs about mathematical discourse; such a measure will provide teacher educators with the tool to understand teacher beliefs and develop programs or interventions that lead to positive change in their beliefs. Chapter 3 reviews the process of developing a survey drawing from the strengths of two frameworks discussed in this chapter. Chapter 4 reviews the research design, which includes recruitment of study participants, their demographic characteristics, summary of the survey items, and data analysis plan. Chapter 5 presents empirical evidence, followed by discussion and conclusion in Chapter 6.

CHAPTER 3. DEVELOPMENT OF THE INSTRUMENT

This chapter includes details about the four building blocks used in the iterative validation process of establishing the validity and reliability of a measurement instrument (Wilson, 2005). The four building blocks were implemented sequentially throughout nine phases. Table 2 summarizes the process. Each data collection point in Phases II and III was followed by analysis that informed the subsequent phase. The theory behind the measurement model is introduced in this chapter, and data on the technical quality of the instrument is represented in Chapter 5.

Cycle	Phase	Building Block	Activity	Chapter
1	I	Construct map	Literature review	3
	II	Item development	Expert paneling of items	
	III		Piloting the BMD items (Version 1) & conducting post hoc think-aloud interviews	
2	IV		Piloting the BMD items (Version 2) & conducting think-aloud interviews (exit interviews)	
	V	Outcome space	Establishing scoring procedures	
	VI	Measurement model	Analysing the data collected using the BMD items (Version 2)	5
3	VII	Item development	Recruitment of study participants Piloting the BMD items (Version 3)	4
	VIII	Outcome space	Revising scoring procedures Training additional rater & scoring of the open-ended responses by two raters	3
	IX	Measurement model	Collecting post hoc think-aloud data from the study participants from Phase VII Analysing the data collected using the BMD items (Version 3) and reporting	5

Table 2 *The Iterative Validation Process*

Phase I: First Building Block: Construct Map

The first step in establishing the validity and reliability of an instrument is the development of a “construct map” (Wilson, 2005). A construct, such as intelligence, efficacy, or belief, is an underlying latent trait that cannot be directly observed and measured. An instrument such as an achievement test or survey is developed and used to uncover something that is latent and can be measured in a consistent way through people’s responses. The construct has a particular form that extends from one extreme to another, from high to low, more to less, small to large, positive to negative, or strong to weak. Thus the construct is continuous. The continuum of the construct is described in the construct map. The purpose of measurement, and the role of the instrument, is to identify where a respondent stands on the continuum that can be defined and described with qualitatively distinct levels. Thus the construct map can be also considered as the theoretical framework that guides the instrument design.

Defining the construct of interest is imperative to the process of measuring it. In this study, the construct of interest is teacher beliefs about mathematical discourse, in short BMD. Mathematical discourse is defined as the way that teachers and students use language to discuss mathematics in the classroom. Teacher belief about mathematical discourse is defined as what a teacher believes is the right course of action in leading a mathematical discussion in a whole-class format[1]. Research shows that teachers hold a variety of beliefs about mathematics (Handal, 2003). In this study, the continuum of teacher beliefs is outlined by the construct map, which covers a range of teacher beliefs about how mathematical discourse should be used in elementary classroom.

The theoretical framework of Scott and Mortimer (2005) was used to develop the construct theory that guided the instrument design, organized on qualitatively distinct levels of teacher beliefs about mathematical discourse. Central to this theoretical framework is the concept of a communicative approach that exists in two dimensions: the dialogic-authoritative dimension and the interactive-noninteractive dimension. The four classes of a communicative approach are defined in terms of whether the classroom communication is authoritative or dialogic in nature and whether it is interactive or noninteractive (cf. Table 1, p. 13). Thus the concept of a communicative approach can be considered as a multidimensional construct. This theory is more fully discussed in Chapter 2.

However, this theoretical framework with intersecting dimensions was not originally conceptualized in terms of a continuum as required for the construct, even though it can be thought of as multidimensional, describing both a teacher's interactive approach and dialogic approach. Using this framework as a basis, the construct map was developed as a continuum to explore and describe qualitatively distinct levels of teacher beliefs about mathematical discourse namely, Univocal, Partial Univocal, Emerging Dialogical, and Dialogical (Figure 1). These levels are described in detail in the following sections.

Belief in Whole-Class Discussion

Respondent	Responses
	More Belief ↑
Dialogical	It is important for the teacher to engage students in conversation with one another to justify their mathematical thinking and argue about mathematical ideas.
Emerging dialogical	It is important for the teacher to encourage students to explore multiple ways of thinking about mathematics and to share these thoughts with one another.
Partial univocal	It is important for the teacher to ask questions to check students' mathematical reasoning and to guide them to the correct answer.
Univocal	It is important for the teacher to explain mathematical ideas and to show how to solve mathematical problems using a step-by-step process.
	↓ Less Belief

Figure 1 *Construct Map*

This construct map allows the representation of a range of teacher beliefs about mathematical discourse. For example, a teacher at the higher end of the construct map (Dialogical discourse) may believe that the primary use or purpose of communication is to encourage student discussions with each other in order for them to explore mathematical ideas, find multiple ways to solve problems, and clarify and challenge each other's ideas in a whole-

[1] Mathematical discourse in the context of students working in small groups or in pairs is excluded from this definition.

class setting. On the other hand, a teacher at the lower end of the construct map (Univocal discourse) may believe that the primary use or purpose of communication is for the teacher to present clear explanations of mathematical ideas. The arrow represents the range of beliefs about mathematical discourse between the two extremes. One may assume that one end of the scale (Dialogical discourse) is better than another (Univocal Discourse) because they are hierachical in nature. However, the relationship between teachers' beliefs about mathematical discourse and the outcomes in terms of students' achievement and understanding in mathematics has not been explored yet (Brendefur & Frykholm, 2000).

Defining the levels.

The actual construct map, with descriptions of each level, appears in Appendix A. The lowest level, Univocal, describes teachers who explain specific ways of reasoning to students (Frykholm, 1999; Scott & Mortimer, 2005). These teachers focus on conveying the exact meaning of mathematics concepts and procedures by showing and explaining the correct methods to solve the problems. Teachers at this level focus on the correctness of student answers; they neither encourage students to articulate their methods nor support students in engaging in conversation with one another in a whole-class format. The Univocal level characterizes a group of teachers who believe that the primary use of communication in the classroom is for them to explain a correct procedure to students. They also believe that the teacher should point out the incorrectness of a student's answer with correct explanations, and that students' explanations may confuse other students since explanations tend to be incomplete and difficult to understand (Chapin, O'Connor, & Anderson, 2003).

The level above the Univocal level, Partial Univocal describes teachers who explain specific ways of reasoning to students, but use communication as a way to check students' understanding of these specific ways of reasoning. Teachers at this level focus on conveying the exact meaning of mathematics concepts by explaining the correct methods to solve the problems. In the process, teachers tend to lead students, either individually (teacher-student interaction) or collectively (teacher-class interaction), through an Initiation-Response-Evaluation (IRE) or Initiation-Response-Feedback (IRF) process. Questions used by teachers at this level are more corrective in nature than explorative. Classroom discourse at this level is dominated by teacher talk and teacher control of questioning (Bearne, 1999). Teachers at this level neither encourage nor support students to engage in conversation with one another in a whole-class format. Thusteachers at this level do not facilitate student-to-student conversation in a whole-class discussion. Teachers in this level may provide students an opportunity to talk to each other either in pairs or small groups. This level characterizes a group of teachers who believe that it is important to explain step- by-step processes for students to arrive at correct answers.

The next highest level, Emerging Dialogical, describes teachers who elicit the individual reasoning of as many students as possible (Scott & Mortimer, 2005). The communication at this level focuses on interactions among students and between teacher and students in which the conversation is limited to assistance or sharing (Cobb et al., 1997). These teachers focus on supporting students to articulate their thinking, but students' articulations are mostly directed to the teacher rather than the class as a whole. Teachers at this level rarely facilitate student-to-student discussion in a whole-class format. An Emerging Dialogical level characterizes a group of teachers who believe that it is important for the teacher to probe an individual student's thinking, but with little awareness about the importance of student-to-student discussion in the whole-class format.

The highest level, Dialogical, describes teachers who explore a range of students' reasoning, with the active participation of students, for the construction of mathematical knowledge and the subsequent shaping of instruction (Scott & Mortimer, 2005; Steffe & D'Ambrosio, 1995). Teachers at this level focus on supporting individual students to articulate their mathematical ideas. Moreover, these teachers focus on engaging students in conversation with one another to share solution strategies and to argue about mathematical ideas in a whole-class discussion. This level characterizes a group of teachers who believe that it is important for the teacher to facilitate the discussion in such a way that students explore mathematical ideas and share their ideas with as clear articulation as possible while also interacting with their peers as a way to learn mathematics.

The construct map, as described above, is necessary for establishing the content validity of a measure. According to the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999), content validity is defined as the degree to which an instrument reflects the specific intended domain of content (Carmines & Zeller, 1991). In this study, content validity was established by literature reviews and item paneling to confirm the relevance of the items being tested, which will be discussed in the following section.

Phase II: Second Building Block: Item Development

The second building block of instrument development is item design (Wilson, 2005). This addresses content validity, which primarily concerns whether an instrument measures the content domain it is designed for. Content validity is also about the degree to which the items adequately and representatively sample the content area to be measured.

The goal of the chosen items is to measure teacher beliefs about mathematics discourse. Since this construct cannot be directly observed and measured, the goal is to assess the construct indirectly. With this in mind, the BMD items were designed and reviewed in various iterations over six months from July to December 2009. Expert judgment was the primary method used to determine content validity of the measure. This section describes the process of the item design to show how research-based literature, the expert panel, and think-aloud interview data informed the development of the final version of the BMD items.

Initial item design.

A broad mix of item formats including open-ended items, multiple-choice items, and Likert-scale items were explored with the expert panel. The panel of experts included teacher educators in the Graduate School of Education and Developemental Teacher Education (DTE) program at the University of California at Berkeley and California State University, East Bay; math educators in the Bay Area Math Project at Lawrence Hall of Science; and a master teacher in Cal Teach program at the University of California, Berkeley. Because of the limitations of Likert-scale surveys discussed in Chapter 2, the panel's recommendation was to eliminate Likert-scale items as a possibility. Inspired by the items designed by Ball and her colleagues (Ball et al., 2004, 2005) to measure teachers' Mathematical Knowledge for Teaching (MKT), I designed the BMD items as scenarios focused on students' correct and incorrect mathematics reasoning as a way to elicit what teachers believe about mathematics discussion and its role in enabling the mathematical learning of and by students. The review of the literature informed the development of the items in terms of the content and type.

As the first step in constructing any assessment, a "domain map," or a description of the topics and knowledge to be measured, should be set out (Ball et al, 2004). The BMD assessment is intended to measure teachers' beliefs about mathematical discourse in elementary classrooms, and therefore I needed to select particular mathematical domains as the focus of discussions. The panel of experts agreed that Number and Operations should be the mathematical domains because these domains are especially important for elementary teaching, dominate the school curriculum, and are vital to students' learning (Ball, Hill, & Bass, 2005; NCTM, 2000; National Research Council, 2002). The Number strand includes whole numbers, decimals, and fractions. For whole numbers, this includes properties of numbers, such as odd and even, prime and composite, square numbers, and factors, multiples, and divisibility. For fractions, it includes the meaning of a fraction, equivalent fractions, and simplifying fractions. The Number strand also includes exploring the relationship between fractions and decimals (NCTM, 2000). The Operations strand includes addition, subtraction, multiplication, and division. This includes understanding meanings of these operations, understanding and developing competency with basic facts, multi-digit computation with whole numbers, and any computation with decimals or fractions or integers (NCTM, 2000).

According to the National Council of Teachers Mathematics Standards (NTCM, 1998, 2000), all students in grades preK-2 should develop a sense of whole numbers, and in grades 3-5, all students should understand the place-value structure of the base-ten number system develop fluency in adding, subtracting, multiplying, and dividing whole numbers; and understand various meanings of multiplication and division.

The first version of the BMD items included a total of nine items, which appear in Appendix B. As illustrated in Figure 2, each item began with a scenario in which a student solves a math problem and the respondent is asked to choose a method for guiding discussion of the student's reasoning ("The teacher should..."). Students' reasoning, both correct and incorrect, about mathematical concepts and procedures was adapted from Chapin, O'Connor, and Anderson (2003). The contextual information was provided to indicate when a whole-class discussion was designed to take place, typically right after a small group activity or working in pairs as a way to reconvene the discussion with every student in the classroom.

Wilson (2005) recommends that the best way to construct a fixed set of responses is to construct an equivalent open-ended outcome space and choose representative responses as the fixed responses. In another words, it is better to start with collecting open-ended responses, and then look for patterns and popular answers in order to construct multiple choice items. However, multiple-choice items were explored in the initial stage for the following reasons. First, multiple-choice items are less time-consuming for respondents, which is important since survey completion time was a serious challenge for recruitment and participation in the study. Second, there was uncertainty regarding whether open-ended items might constrain respondents' ability to articulate their responses. Thus the response options for each multiple-choice item were developed to reflect characteristics of each level of the construct map so that there was a direct relationship between each response option and each level of the construct map. Figure 2 is one example of a multiple-choice item. As will be discussed in the next section, the initial responses to the multiple-choice answers were used as a pilot version of the instrument and then queried in interviews. These interview questions were key to the ongoing development and improvement of the instrument.

Directions: For each item, choose ONE answer that you feel is the MOST important.

[Context 1: After the discussion of the problem on the board, the teacher gives another multiplication problem (15 X 17) to the students to work in pairs. After students have worked on the problem in pairs, the teacher reconvenes the whole class for discussion.]

1. The teacher calls on a student to present his computation to the whole class and his solution is incorrect.
On the board:

```
  15
X 17
 105
+ 15
 120
```

In this whole-class discussion, the teacher should:

a. Model the correct method.
b. Ask questions to help him think through his reasoning.
c. Correct the student to help him understand the correct method.
d. Ask other students to comment and help the student think through his reasoning.
e. Other: ____

Are this question and response options clear to you? Yes ____ No _____
If not, please explain what is unclear to you:

Figure 2 *Sample Multiple-Choice Item*

Phase III: Piloting the BMD Items (Version 1) & Think-Aloud Interviews

The first round of item piloting was conducted with second-year preservice teachers (N=20) enrolled in the Elementary Mathematics Method Course in the Developmental Teacher Education (DTE) program at the University of California, Berkeley in September, 2009. The first version of the BMD instrument was in pencil and paper format and administered in the classroom at the end of the course. Participation was voluntary, and the instrument took 15-20 minutes to complete. No demographic data were collected at that time. The item piloting showed that the design needed to undergo a change in terms of item type and the difficulty of the mathematics problems used in the examples.

Response option	a. Model the correct method.	b. Ask questions to help him think through his reasoning.	c. Correct the student to help him understand the correct method.	d. Ask other students to comment and help the student think through his reasoning
Frequency (n)	1	4	6	9

Table 4 *Frequency of Item #1*

The frequency distribution of teachers' responses for all nine items showed that the data were highly skewed such that a majority of preservice teachers fell into the higher levels in the continuum, namely level 3 (Emerging Dialogical) and level 4 (Dialogical). It is possible that these preservice teachers genuinely believed that mathematical discourse was important to support students to explore mathematical thinking and to engage students in a discussion with other students in a whole class discussion. However, these skewed numbers may also have resulted from the problem of self-reporting.

After the first round of the item piloting, two respondents were recruited to participate in the post-hoc think-aloud interviews. The post-hoc think-aloud interview protocol appears in Appendix C. The think-aloud interview was conducted with each preservice teacher individually for thirty minutes. The think-aloud interview (Ericsson & Simon, 1993) was used to explore the thinking of the respondents as they worked through an item. It allowed them to verbalize their decision making and rationale behind each response choice. The respondents were asked to share their thought processes and identify any language in the items that was unclear and to point out any items that were particularly difficult to respond to in the survey. The post-hoc think-aloud interviews revealed that the respondents were able to identify the response options that were mapped onto the higher levels in the continuum, even when these options were inconsistent with their beliefs. For example, a respondent said,

> I knew right away that the option (a) wasn't the right answer. As a teacher I believe the teacher should ask the student about his solution because the teacher should understand what the student was thinking. But I knew (d) is what I was supposed to choose, and I did pick (d) because we in our Math Method Course talk a lot about giving the students the opportunity to learn from one another by asking each other questions and helping each other during that process. However, I personally don't think that kids can manage that.

Based on the interview data, the interviewee's initial survey answer can be described as socially desirable responding (SDR) (Mick, 1996), which is the tendency of respondents to describe themselves in favorable terms by adhering to socio-culturally upheld norms. SDR can adversely affect the validity of studies in many social science disciplines (De Jong, Pieters, & Fox, 2009). The post hoc think-aloud interview was helpful to gather validity evidence of the instrument.

Both respondents also expressed that all of the response options appeared to be reasonable, but that their choice would vary depending on their instructional goals. Without sufficient contextual information, the respondents determined that what the teacher should do was to lead a discussion. The respondents also expressed that they found choosing just one answer difficult, and both respondents chose multiple response options for a single item. In some situations, the highest level of the construct was not necessarily the best course of action to take to lead a mathematical discussion. I recognized the need to add a question to probe the reasoning behind answer choices to provide more robust evidence of the validity of the BMD items. For example, when the teacher introduces any mathematical term (item #8, shown in Figure 3), the respondents expressed that the teacher has to start with a clear definition. Although the response option (c) was mapped into the highest level in the continuum, the respondents strongly believed that asking students to share their definitions was not the most pedagogically appropriate action to take.

The teacher is introducing the mathematical term "even, and odd" when talking about numbers. For this whole class discussion, the teacher should:

a. Provide students with a definition.
b. Chart students' ideas as students share their definitions.
c. Ask students to share their definitions and invite others to comment.
d. Provide students with a definition of "even" and then ask them to guess what "odd" might be.
e. Other: ______________________________

Figure 3. Item #8

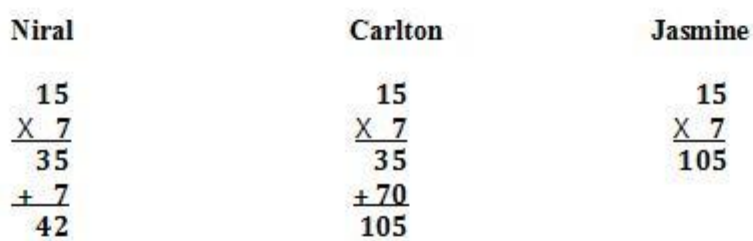
After the think-aloud interviews, item #8 was eliminated. Finally, a respondent expressed that the items overall were not applicable to teachers who may want to teach lower level grades (e.g., K-2) because the mathematics problem used in the most items were more applicable to grades 3 to 5.

Item revision based on item piloting and think-aloud interviews.

The first round of item piloting and the post hoc think-aloud interviews with two participants informed the revision of the BMD items. The following changes were made. First, a majority of the BMD items, all but three, were changed to open-ended questions in order to reduce the possibility of the respondents recognizing the answer that would map into higher levels of the construct. Second, an additional question, "Explain why your choice is most important" was added to the three multiple-choice items to provide the respondents with an opportunity to explain their professed course of action, which was an importance piece of information to assess teacher beliefs about mathematical discourse. Third, the items were revised to evenly cover grades K through 5, with various levels of mathematics content difficulty. To cover items relevant to each grade, some mathematics problems used in the items were revised to be simpler. For example, an item about multiplication of a two-digit number by a two-digit number (e.g., 15 X 17) in the initial version of the BMD assessment was changed to multiplication of a two-digit number by a one-digit number (e.g., 15 X 7) to help respondents relate to the items more easily. Fourth, the item stem of "In this whole class discussion, the teacher should…" was changed to "In this whole class discussion, what should the teacher SAY and DO to …." Since the goal of the BMD items was to understand teachers' beliefs about discourse specifically, it was important to elicit the responses that were directly related to the teacher's approach to a mathematics discussion. Thus the item stem was changed to prompt teachers to reveal what they should say or do in a given situation.

The revised version of the BMD assessment included ten items, which consisted of seven open-ended items (item #1 through #7) and three multiple-choice (item #8 through #10) with justification items. Figure 4 shows an example of the open-ended items.

4. A third grade teacher asks students to work individually on the problem 7 x 15. The teacher reconvenes the class and calls on three students to present their computations on the board.



How should the teacher guide a discussion about the three computations?

Figure 4 *Example of Revised Open-Ended Item*

For three multiple-choice with justification items, respondents were asked to choose the best option for the teacher given the situation and explain why the chosen option was the most important action for the teacher to take (Figure 5).

8. The mathematical term “multiples” was introduced to students in third grade. This year, their fourth grade teacher asks students to work individually to find multiples among numbers on the board:

8, 12, 16, 20, 24, 28, 32, 36

When the teacher reconvenes the class, what is MOST important for the teacher to do with the class?

a. Ask for answers, and highlight correct answers while reviewing the definition of ‘multiple.’
b. Ask students to agree or disagree with one another’s ideas and explanations.
c. Guide students to identify multiples of 4, and then ask whether all numbers are also multiples of 8.
d. Elicit several students’ ideas about multiples, asking questions to check for understanding.
e. Other (please specify)

Explain why your choice is most important:

Figure 5 *Example of Multiple-Choice with Justification Item*

At the end of Phase III, the BMD items were revised given respondent feedback and further research. The revision was undertaken with the goals of strengthening the clarity of the questions and improving the validity and reliability of the instrument. Version 2 of the BMD is discussed in the next section.

Phase IV: Piloting the BMD Items (Version 2) & Think-Aloud Interviews

In the second round of item piloting, the revised BMD items were administered to 35 preservice teachers who were enrolled in the Elementary Mathematics Method Course in teacher education programs in California State University, East Bay in November, 2009 during the winter quarter. The second version of the BMD measure is found in Appendix D. The study participants were a group of preservice teachers with diverse academic backgrounds, which include art history, business, creative writing, sociology, psychology, political science, theatre, liberal study, speech communication, and kinesiology. Table 5 shows the demographic characteristics of participants in the study.

Characteristic	*N*	*Percentage*
Gender		
Female	25	83%
Age		
Under 25	5	17%
25-29	9	30%
30-39	8	27%
40-49	4	13%
50-59	4	13%
Ethnicity		
Asian	4	13%
African American	3	10%
Caucasian	19	63%
Hispanic	3	13%
Other	3	10%
Total	30	

Table 5 *Demographic Characteristics of BMD Version 2 Teachers*

The participants were in their third quarter of the four quarter program. They began during summer quarter in 2009 and had their first placement in the fall quarter of the same year. At the time of item piloting, they were starting their second field placement. Of the study participants, 72% reported that they would like to teach in a suburban area, whereas 55% of the participants desired an urban or inner-city school, and 14% wanted to teach in a school in a rural area after completing the program. A respondent was allowed to choose more than one area that they would like to teach.

The second version of the BMD items included the revised versions of the questions with five open-ended items ("In this whole class discussion, what should the teacher SAY and DO to ….") and five multiple choice with justification items ("Explain why your choice is most important:") where the respondents were asked to choose one of the five answer choices and then explain why they chose that answer.

This version of the BMD instrument functioned better in terms of eliciting teachers' beliefs about mathematical discourse. For example, in the open-ended items, teachers wrote detailed descriptions of the kind of questions they would ask their students and how they would respond to the situation. In the multiple-choice with justification items, teachers wrote detailed explanations that supported their response options. Typical responses of teachers at level 1, Univocal, were that the teacher should point out a student's mistake and show how to solve the problem so that the class can see how to get a correct answer. Teachers at this level expressed that it was the best for the teacher to explain how to solve the problem because students' explanations might easily confuse their peers. Typical responses of teachers at level 2, Partial Univocal, expressed the importance of understanding a single student's thinking and asking questions to guide them to the correct answer. The discussion was mainly between the teacher and the focal student. Typical responses of teachers at level 3, Emerging Dialogical, emphasized the importance of the students sharing their answers with one another so that students know that there are multiple ways of solving mathematics problems. Typical responses of teachers at level 4, Dialogical, expressed that it was important for the teacher to give students the opportunity to share their answers and ask each other questions so that they can help one another and use one another as the resources in learning mathematics.

After the second round of item piloting, two respondents volunteered to participate in post hoc think-aloud interviews, which yielded useful information about the items for improvement. For example, one interviewee said that it was difficult to answer item #5 since no student work was provided, unlike other items. The interviewee said that she needed to see the student's answer to determine how to lead a discussion. The other interviewee pointed out that, for item #9, she knew that a good teacher would not choose the first response option (to tell the class which set-up is correct). She also added that the remaining response options were all appropriate and she had difficulty choosing one. The same comment was made for item #7. The multiple-choice options for items #6, #8, and #10 worked well to elicit teachers' explanations. The feedback was incorporated to the revision of the BMD measure. Appendix E shows the changes made from version 2 to version 3 of the BMD measure. The final version of the BMD measure is presented in Chapter 4.

Phase V: Third Building Block: Developing a Scoring Guide

Once the responses to the items were gathered, they were categorized and scored to be indicators of the construct, which led to the third building block of the instrument design, the "outcome space" (Wilson, 2005). The categories that define the outcome space are qualitatively distinct. The outcome space maps directly onto its corresponding contruct. Data from this phase of instrument design provide evidence for the instrument's content (Research Question 1)[2] and allow for inter-rater reliability study (Research Question 5)[3]. Two different sets of scoring guides were developed, one for open-ended items and one for multiple choice items. These scoring guides provide a framework for the interpretation of the responses to reveal what the respondents believed about mathematical discourse in the elementary classroom. This section describes the process of categorizing and scoring different responses to the open-ended items and multiple-choice with justification items with the scoring guide as the end result.

A scoring guide, Table 6, shows the levels of the construct map, example responses, and corresponding scores at each level. The guide categorizes the data so that the data can be related back to the construct using the measurement model. Table 6 is a generic example that summarizes the outcome space of open-ended items #1 through #7. For these items, the outcome space is ordered into qualitatively distinct and ordinal categories. Responses reflecting teachers' Univocal communication receive the lowest score whereas responses reflecting teachers' Dialogical communication receive the highest score, which is consistent with the conceptualization of teacher beliefs from Univocal to Dialogical communication as described by the construct in Figure 1.

Since items #1 through #7 are open-ended items, the responses were scored by two scorers. According to the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999), inter-rater reliability is the extent to which two or more individuals (scorers or raters) agree. Inter-rater reliability addresses the consistency of the implementation of a rating system (Research Question 5: What is the evidence for reliability?). A scoring training session was conducted in January and Feburary 2010 to provide the other rater (the principle invesigator was the primary rater) with information about the items in relation to the construct map and to train the rater in how to score the instrument. I, the principal investigator, and the other rater reviewed the items in relation to the construct map and item response examplars, and scored a wide range of responses that were mapped into levels of response. Table 6 is a generic scoring guide. An item-specific scoring guide appears in Appendix F. Since all item-specific scoring guides were developed from the same construct map, they could also be interpreted similarly across items. The scoring training session helped to create robust scoring guides unique to each item. Five scoring moderation sessions took place between January and February 2010 to train one additional researcher to score responses on the open-ended items. Two researchers double scored twenty percent of the data to determine the degree to which they agreed in the scoring of the items. A total of 40 surveys were double scored by two raters. The inter-rater reliability is reported in Chapter 5.

---

[2] To what degree, if any, is teacher beliefs about math discourse associated with other factors such as demographics, educational background and teaching experience?

[3] What is the evidence for reliability?

Level on the Construct Map	Description	Score
Dialogical	The role of communication is as primary way for students to engage in mathematical ideas with one another in a whole-class discussion as reflected in their language: - Encourage students to agree or disagree with other students' thinking and explain why - Encourage students to ask questions, respond to, or make comments on other students' thinking Example: "*The teacher should have the students challenge their own mathematical thinking and other's thinking, including multiple ways of solving the problem by asking questions and making comments*."	4
Emerging Dialogical	The role of communication is for teacher to engage students to explore multiple ways of thinking of mathematics, and encourage students to hear other students' mathematical thinking as reflected in their language: - Encourage students to explore multiple ways of solving mathematical problems - Encourage students to share and hear multiple ways of solving mathematical problems with one another Example: "*The teacher should encourage students to share different ways of problem solving*."	3
Partial Univocal	The role of communication is for the *teacher* to deliver specific ways of mathematical reasoning and check correctness of students' answer as reflected in their language: - Use communication to convey the exact meaning of math concepts and as a way to lead students to a correct answer and method, using a series of questions, hints or step by step process - Communication is mainly used in a way for teacher, not the whole class, to hear what a student says Example: "*The teacher should walk a student through a problem to show the student the correct steps*."	2
Univocal	The role of communication is for the *teacher* to deliver specific ways of mathematical reasoning, without attending to students' mathematical reasoning as well as their opportunity to verbalize their mathematical reasoning, as reflected in their language: - Use communication to explain mathematical concepts and/or show the method - Use communication to point out the incorrectness of students' responses and explain why Example: "*The teacher should point out that an answer is incorrect when a student gives an incorrect answer, and explain what was wrong with an answer*."	1
Irrelevant response	Teacher mentions aspects of math instruction that are irrelevant to the discourse, such as small group work or use of manipulatives. Example: "*The teacher should provide students with manipulatives*."	0
Missing	Missing response.	9

Table 6 *Scoring guide for open-ended items*

Phase VI: Fourth Building Block: Measurement Model

Once responses to items were scored using the outcome space, the scored responses were related back to the construct map. This is the fourth building block, the measurement model (Wilson, 2005). This study uses Item Response Theory (IRT) to examine the relationship between the empirical results and the construct map. This section describes two competing measurement models to show the benefits of using the IRT model. IRT models have not been utilized in research on teacher beliefs in the past, and the development of an instrument to measure teacher beliefs about mathematical discourse using IRT will be a novel contribution to the field of research on teacher education.

Classical Test Theory (CTT) vs. Item Response Theory (IRT).

There are two competing test theories in educational measurement: Classical Test Theory (CTT) and Item Response Theory (IRT). A brief discussion of the two test theories is approprate to explain why IRT is used in this study.

CTT assumes that the raw score (X) obtained by an individual is made up of a true score (T) and some unobservable measurement error (E) (Hambleton & Swaminathan, 1991).

$$X = T + E \qquad (1)$$

However, the true score derived from CTT is not an absolute characteristic of the respondents. For example, consider a group of students who are given two math tests. First, a student's score may be higher on one test than the other because of the difficulty of the tests. On the other hand, a math problem may be considered to be difficult for low-performing students whereas the same problem may be considered easy for high-performing students. That is, how easy or difficult the test is also depends upon who is taking the math test. According to CTT, item difficulty is group dependent, and is different for every group. This applies to a case in which an individual performs differently on two similar tests depending on several factors, such as the psychological (e.g., distracted or depressed) and physical state of an individual (e.g., sleep deprived) and the test environment (e.g., noisy). In the CTT model, there are several assumptions to fulfill several purposes on statistics; (a) random errors in CTT are normally distributed with a 0 expected value and not correlated with each other, and (b) it assumes equal errors of measurement for all examinees, and (c) random errors are uncorrelated with each other. Since those random errors are uncorrelated with each other; there is no systematic pattern to why scores would fluctuate from time to time (Kline, 2005).

Since the total score is only as good as the sum of its parts (items), item level investigation is important. That is why IRT is used in this study. IRT correlates test and item scores based on assumptions concerning the mathematical relationship between abilities (or other traits, such as attitude) and item responses (Hambleton, Swaminathan, & Rogers, 1991). IRT models are functions relating person ability and item difficulty parameters to the probability of a discrete outcome (e.g., correctness or agreement). For example, a high ability respondent would be predicted to have a higher probability of answering an item correctly than a low ability respondent on any particular item.

IRT measurement models, when compared to CTT models, offer several distinct benefits. First, item statistics are independent of the sample from which they were estimated. This means that the difficulties of items are not specific to the group of respondents who take a test. Second, examinee scores (raw scores) are still dependent on test difficulty but their estimated ability is independent of test difficulty.

Rasch measurement.

The Rasch model is expressed both at the item level and the instrument level. The Rasch model focuses on modeling the *probability* of the observed responses, rather than on modeling the responses themselves (Wilson, 2005). In achievement application, the "respondent location" is thought of as the respondent ability whereas in attitude or belief application, the respondent location can be understood as an attitude or belief towards something. The "item location" is the item difficulty or item scale value. In this study, the terms "respondent location" and "item location" are used. The proability of the item response for item i, X_i, is modeled as a function of the respondent location (θ) and the item location (δi). The probability of success (e.g., attitude or belief toward something) is a function of the difference between the location of the respondent and location of the item. The logic of the Rasch model is that the respondent has a certain amount of the construct, θ, and an item also has a certain amount of δ. The difference between the person and item locations determines the probability. Putting this in the form of an equation, the probability of response $X_i = 1$ is:

$$\text{Probability}(X_i = 1 \mid \theta, \delta_i) = \frac{e^{(\theta - \delta_i)}}{1 + e^{(\theta - \delta_i)}}. \qquad (2)$$

The empirically calibrated version of the construct map is the Wright Map (Wright & Masters, 1982), which is a tool for a visual representation of data that is used to check whether the empirical results support the internal structure of the construct that is hypothesized (Wilson, 2004). Hence, on the Wright map, the difference between a respondent's location and the corresponding item location governs the probability that the respondent will make a particular response. The BMD measure was calibrated using ConQuest (1998), which is a estimation software for a generalized Rasch item response model developed by Adams, Wilson, & Wang (Wu, Adams, & Wilson, 1998). ConQuest also provides person separation reliabilities (Research Question 5).

Summary

In this chapter the development of the Beliefs on Mathematical Discourse instrument has been discussed. The various phases of the process were reviewed with an eye towards the collection of vital data for the production of a final version of the instrument. Clearly, multiple data sources, literature, expert input, and individual feedback from respondents, were all crucial to refining the development of a reliable and robust tool. In the next chapter, the research design of the study will be presented around the discussion about the participants, additional questions to the BMD measure, and data analysis plan.

CHAPTER 4: RESEARCH DESIGN

This chapter discusses the research design of the study using the final version of the BMD measure (Appendix G). The development of the BMD measure was described in Chapter 3. This chapter has three sections. The first section is about the participants who were recruited to take the final version of the BMD measure. The second section presents additional questions that were asked of the participants to answer the research questions. The third section includes the data analysis plan.

Phase VII: Research Design and Item Development

Participants.

This section presents the study population and the sampling procedure used. The demographic, educational, and teaching backgrounds of participants in this sample are presented.

In this study, the survey population is defined as individuals who are currently teaching in elementary school classrooms or currently enrolled in teacher education programs with a plan of teaching in elementary school classrooms in the near future. The study does not draw from a nationally representative sample across teacher education programs and elementary schools in the U.S. My purposive sampling was a decision for a recruitment purpose, and also more informative than random sample for the purpose of developing an instrument.

Three sample pools were developed to purposely select participants from the population. The first pool includes individuals who were enrolled in teacher educaton programs in the Spring semester of 2010. The second pool includes individuals who graduated from teacher education program in Fall 2009. The third pool includes individuals who are currently teaching full-time in elementary classrooms. The individuals in the population vary in terms of their backgrounds in mathematics, their amount of exposure to teacher training including math methods courses and field placement, and their years of experience teaching elementary mathematics.

Recruitment.

The May 2004 compact between Governor Schwarzenegger and California's higher education community identified filling the critical shortage of K-12 mathematics and science teachers as a major state priority (Guha, Campbell, Humphrey, Shields, Tiffany-Morales, & Wechsler, 2006). A commitment was made by the California State University system to double the number of mathematics and science teachers it trained by the year 2010 through the Mathematics and Science Teacher Initiative. The California State University system plays an important role in recruiting and training mathematics teachers for the nation's schools since it trains so many teachers compared to the the University of California, which has not been traditionally considered a significant contributor to the overall teacher supply. Thus both the California State University system and the University of California were crucial in recruiting preservice teachers for the study.

To recruit the sample pool of preservice teachers who were enrolled in teacher educaton programs at the time of the study, instructors teaching math methods courses in elementary teacher education credential programs were contacted via e-mail. The e-mail provided the university instructor with an explanation of the study and requested that the researcher be allowed to use their class as a source for preservice teachers who would be interested in taking the survey. Teacher educators who agreed to participate in the study disseminated the online survey to their students via e-mail. However, it was the preservice teachers who ultimately decided whether they wanted to participate in the survey, and they were assured that the instructor would not be informed regarding participation. Since the online survey link was sent to teacher educators for dissemination, names and e-mail addresses of the preservice teachers were unavailable to the principal investigator.

The second sample pool included individuals who had recently graduated from teacher education programs, but might not yet have entered teaching at the time of data collection. A director of one teacher education program in the San Francisco area disseminated the online survey to students who graduated in Fall 2009. The third sample pool consisted of teachers teaching in elementary school classrooms. The online survey was posted on a New York-based elementary teacher's Facebook page, a social networking site, for recruitment. Table 7 presents pertinent demographics for the sample pool.

# of completed responses	# of the respondents who started the survey but did not complete	Total
167 (82%)	35 (17%)	202 (100%)

Table 7 *Breakdown of the Sample*

Description of sample characteristics.

The data using the final instrument were collected from 202 respondents from the beginning of January to the beginning of Feburary, 2010. Out of 202 respondents, 17% did not continue the survey after one or two questions. These respondents were dropped from the study. Table 8 shows the demographic characteristics of participants in the study. The survey was administered using Survey Monkey. A unique person identification number was assigned to each response to assure anonymity and confidentiality.

	All	California State Universities	University of California	Private University[4]	In-service Teachers
Number of Schools	10	4	5	1	n/a
Number of Participants	167	70 (41%)	57 (33%)	13 (7%)	27
Ethnicity					
Asian	11 (7%)	6 (9%)	5 (9%)	-	-
Black	4 (2%)	1 (1%)	2 (9%)	-	1 (4%)
Latino	20 (11%)	12 (16%)	6 (11%)	1 (8%)	1 (4%)
White	111 (67%)	44 (62%)	35 (61%)	9 (75%)	23 (85%)
Other	20 (11%)	7 (9%)	10 (15%)	2 (17%)	1 (4%)
Gender					
Female	148 (90%)	59 (84%)	52 (5%)	10 (83%)	27 (100%)
Male	16 (10%)	11 (16%)	3 (5%)	2 (17%)	-
Age					
Under 25	88 (53%)	40 (58%)	41 (73%)	5 (38%)	2 (7%)
25-29	37 (22%)	12 (17%)	12 (21%)	3 (23%)	10 (36%)
30-39	22 (13%)	8 (12%)	2 (4%)	2 (15%)	10 (36%)
40-49	13 (8%)	7 (10%)	1 (2%)	2 (15%)	3 (11%)
50-59	2 (1%)	2 (3%)	-	-	2 (4%)

Table 8 *Breakdown of the Sample by Program (n=167)*

The sample characteristics of this study reflect mirror the breakdown of California's teaching force. In 2008-09 California's teachers were predominantly white (70.1%) and female (72.4%), quite a different profile from the student population that was 51.4% male and had major ethnic categories of 49.0% Hispanic, 27.9% white, 8.4% Asian, and 7.3% African-American

[4] The private school here is outside California.

(Education Data Partnership, 2010). A complete description of sample characteristics can be found in Appendix H, along with additional information about the participants. For preservice teachers, the details include participants' number of years in a teacher education program, undergraduate major(s), and the grade level and community in which they would like to teach after completing their program. Preservice teachers came from diverse academic majors, and 48% of them were in their first year in teacher education programs. Of the surveyed preservice teachers, 50% wished to teach in a suburban community after obtaining their teaching credential.

For in-service teachers, their background information includes their undergraduate major, the type of teacher education program they attended in terms of length, from which state they obtained their teaching credential, the state and grade level in which they are currently teaching, and the characteristics of their students in terms of family income. In-service teachers who responded to the survey were spread across 16 states[5]. Most teachers obtained their elementary teaching credential from the state in which they were teaching at the time of data collection. Of the respondents, 30% reported having teaching credentials from multiple states in the United States. The percentage teaching in a self-identified suburban school was 47%, while 30% identified as teaching in an inner-city/urban area, and 24% in rural areas. Twenty of the teachers had 5 or more years of teaching experience. Half of the teachers characterized their students as low-income and the other 50% of teachers characterized their students as middle-income. Of the inservice teachers, 50% reported having attended 2-year teacher education programs whereas 31% attended 1-year teacher education programs, and 19% attended alternative certification programs such as Teach For America or Teaching Fellows.

BMD measure.

The final version of the BMD items (Appendix G), a total of 10 items, represents a range of complexity in math content, including strands of Number and Operations to cover from grade K-5 of elementary school mathematics. Item #1 is about patterns. NCTM Standards (1998) contain numerous references to patterns. In grades preK-2, all students should sort and classify objects by different properties; order objects by size or other numerical properties; and identify, analyze, and extend patterns and recognize the same pattern in different manifestations. Reasoning about sequences of attributes reinforces understanding of number and function, and furthermore, enforces a better understanding of logic, both the common-sense logic that students use in every class and the more formal logic they need in higher grades to learn about proof as a philosophical/mathematical ideal. Items #2 through #8, and #10 are about Number and Operations. Item #9 is about geometry. According to NCTM Standards (1998), in grades 3–5, all students should be able to identify, compare, and classify two- and three-dimensional shapes according to their properties and develop vocabulary to describe the attributes, and develop definitions of classes of shapes such as triangles and pyramids. Table 9 summarizes the mathematical content of each item.

Item #	Mathematical Content Used in the Items
1	Patterns
2, 3	Subtraction
4	Multiplication and place values
5	Addition
6	Using base-10 blocks to represent whole numbers
7	Subtraction and place values
8	The concept of multiple
9	Sorting geometric figures based on their properties

[5] The 16 states include AZ, CA, CT, FL, GA, MA, MI, IL, IN, KY, PA, MO, NJ, NY, OH, WI.

10	Finding solutions for pairs of equations that are number sentences	

Table 9 *Mathematical Content of Items*

To help the respondents with completing the BMD measure, questions that were easiest in terms of grade level were placed at the beginning of the instrument. The items were ordered and presented from the simplest math problems to more complex math problems in the survey. The multiple-choice with justification items were placed at the end of the instrument in the hope that the respondents' answers would not be influenced by the content of the response options of the multiple-choice items.

Demographics and convergent validity items.

This section presents additional items that were used to collect data to answer the research questions. A total of 6 items were designed and added to collect respondents' demographic information (e.g., age, gender, ethnicity, etc.) along with academic background. Depending upon respondents' current status, they were required to respond to a different set of these additional questions. For preservice teachers, they were asked about the years of teacher education training they had received at the time of the study, and the kind of community and grade level in which they would like to teach. In contrast, in-service teachers were asked the kind of community and grade level in which they are currently teaching, years of experience, and the type of teacher education program they attended for an elementary credential. These questions were designed to generate data on the relationship between the instrument to other variables, addressing the last research question. (Research Question 6: To what degree, if any, is teacher beliefs about math discourse associated with other factors such as demographics, educational background and teaching experience?).

Evidence of convergent validity.

When two instruments are intended to measure a similar variable, a strong relation between the external variable and the instrument under investigation is expected. The Preservice Teachers' Attitudes about Discourse in the Mathematics Classroom (PADM) measure was developed by Casa, McGivney-Burelle, and DeFranco (2008) to measure preservice teachers' attitudes about discourse in the mathematics classroom. Since the PADM instrument and the BMD measure intend to measure a similar construct about teachers' attitudes and beliefs about mathematics discourse, one expects to see high a correlation between the two (convergent evidence). Data collected using the PADM instrument was used to explore validity evidence based on an external variable (Research Question 4: What is the evidence for external validity?).

The Preservice Teachers' Attitudes about Discourse in the Mathematics Classroom (PADM) measure has 26 items. This measure uses a 5-point Likert scale, ranging from 1 for 'strongly disagree' to 5 for 'strongly agree.' Validity of the PADM measure was examined by the content experts and reliability was examined using a coefficient of consistency, Cronbach's alpha (α). The authors of the PADM measure report that the instrument has three dimensions: promoting mathematical reasoning (α_1 =.85), examining complex mathematical concepts (α_2 =.81), and valuing students' mathematical ideas (α_6 =.85).

Data Analysis

Measurement models serve as a method to relate scored outcomes and to compare the outcome space to the original construct (Wilson, 2005). In Chapter 3, a brief introduction of two competing measurement models were presented, and the rationale for using IRT, specifically the

Rasch model, in this study was discussed. However, the extensions of the Rasch model are discussed in this section because of the need to fit the polytomous responses in this study: the Partial Credit Model (PCM) (Masters, 1982; Wright & Masters, 1981) and the Rating Scale Model (RSM) (Andrich, 1978).

RSM assumes that step locations remain constant from item to item. Thus the rating scale model has more constraints to meet the assumption that the step difficulties are the same from item to item. This means that moving from one level to another level is equally difficult for all items. In contrast, the strength of PCM is that it gives the data a better fit because step locations are allowed to vary from item to item. The "steps" are defined as the points or thresholds within the construct continuum where the transition between two adjacent categories takes place. Step parameters, δ_{ik}, are used to represent ordered response categories (Edwards & Thurstone, 1952; Samejima, 1969) and govern the probability of making the "step" from score *k-1* to score *k*. The multiple-choice items in the BMD measure have four response categories. Thus each item has three steps. The first step is the transition from Univocal (score=0) to Partial Univocal (score=1), the second step is the transition from Partial Univocal (score=1) to Emerging Dialogical (score=2), the third step is the transition from Emerging Dialogical (score=2) to Dialogical (score=3). Step locations vary from item to item in PCM, meaning that it may be more difficult to move from Partial Univocal to Emerging Dialogical, for example. Some steps can be larger for some items than others. Thus PCM was applied to the data in this study as it best fits the data.

The location of person, *n*, is β_n, *k* is the level of item *i*, δ_{ij} (j=1,…., m_i). The item step difficulty associated is with score, *x*, on the m_i step of item *i*. The score *x* is the *count* of the completed item steps. The Partial Credit Model is stated as:

$$\pi_{\mathrm{nix}} = \frac{\exp \sum_{\mathrm{j=0}}^{\mathrm{x}} (\beta_{\mathrm{n}} - \delta_{\mathrm{ij}})}{\sum_{\mathrm{k=0}}^{\mathrm{m_i}} \exp \sum_{\mathrm{j=0}}^{\mathrm{k}} (\beta_{\mathrm{n}} - \delta_{\mathrm{ij}})} \qquad (3)$$

Where δ_{i0}=0 so that exp $\sum_{\mathrm{j=0}}^{0}(\beta_{\mathrm{n}} - \delta_{\mathrm{ij}})$=0 and exp $\sum_{\mathrm{j=0}}^{0}(\beta_{\mathrm{n}} - \delta_{\mathrm{ij}})$=1. This equation states that the probability a person, *n* has of scoring, *x*, on step m_i, is a function of the difference between a person's ability level, β_n, and the item step difficulty of item *i* δ_{ij}.. The numerator contains only the difficulties of these *x* completed steps, δ_{i1} , δ_{i2} , δ_{i3} ,…. δ_{ix} . The higher the value of δ_{ij}, the more difficult a particular step is relative to other steps within an item. The denominator is the sum of all m_i + 1 possible numerator (Andrich, 1978; Wright & Masters, 1982).

Summary

To summarize, after having developed and piloted the BMD instrument through two rounds of revisions, it was administered to 3 pools of volunteer participants. In addition to the survey, each pool was asked a set of pool-specific demongraphic questions. Two measurement models were tried, with PCM being chosen due to its best fit to the data.

CHAPTER 5: EMPIRICAL EVIDENCE

This chapter presents evidence for the validity and reliability of the BMD measure. The first research question about content validity rests on the connections between the building blocks of my dissertation — the construct map, the item design, the scoring guide, and a measurement model — which were presented in Chapter 3. Five of the six research questions will be answered in this chapter. They are the following:

2. What is the evidence for response process validity?
3. What is the evidence for internal structure validity?
4. What is the evidence for external validity?
5. What is the evidence for reliability?
6. To what degree, if any, is teacher beliefs about mathematics discourse associated with their teaching status?

Research Question 2 (RQ 2): What is the Evidence for Response Process Validity?

According to the *Standards,* "theoretical and empirical analyses of the response processes of test takers can provide evidence concerning the fit between the construct and the detailed nature of performance or response actually engaged in by examinees" (AERA, APA, NCME, 1999). Thus respondents' interpretations of items, and their general reaction to the items in terms of difficulty and clarity, are important information to gather as part of validity evidence. The process of the respondents verbalizing the thought processes they used to answer the items helped to ensure that the items were understood by the respondents as they were originally intended, and that there was no major confusion. The data gathered from the response process were used to revise the BMD measure.

Validity evidence based on response process was collected at three subsequent times after piloting each version of the BMD measure throughout the research. Response process from Phase III and IV using the version 1 and 2 of the BMD measure were described in Chapter 3, which informed the development of the third version of the BMD measure. In this section, the results of the response processes during Phase IX using version 3 of the BMD measure are described below to suggest evidence towards the validity of the BMD measure.

During Phase IX, 202 respondents participated in the survey using the online survey tool, Survey Monkey. The post-hoc think-aloud interviews were conducted with five respondents (one current teacher and four preservice teachers) who participated in the survey during this period. The interviewees were asked to go over one item at a time to verbalize their thought processes in giving answers, identifying any questions or any part of the questions that were too challenging to answer or that they did not understand. According to think-aloud interviews, the BMD items were clear to the respondents in terms of the goal and wording of the questions, but there were several limitations with items.

First, different ideas about "discussion" were operating in three respondents' minds when they were answering the survey. One interviewee described discussion mainly as verbal interaction between the teacher and students. Another interviewee agreed with the first interviewee, yet also highlighted the interaction among students asking questions and responding to one another's ideas. In contrast, a third interviewee included non-verbal interaction also as a way to participate in discussion, especially for students in lower grades. For example, when a student presents a solution on the board for other students to see with or without any verbal explanation, in some teachers' minds, that student was participating in discussion because that student was sharing a solution with the rest of the class. However, two respondents emphasized that the most important characteristic of discussion is students agreeing or disagreeing with their peers about their mathematical ideas, not just simply presenting and sharing each other ideas.

Second, it was found that some items on the BMD measure were framed positively in relation to discussion. For example, the question in item 3 says, "What should the teacher say and do to guide students toward an understanding of subtraction?" Two respondents pointed out that the word "guide" was a signal that discouraged them from their initial response that the teacher should lecture about subtraction. Likewise, the question in item 4 says, "How should the teacher guide a discussion about the three computations?" One interviewee pointed out that this question was framed in a way that values multiple students' mathematical ideas, which is more positive toward discussion. She said she initially did not even think about eliciting multiple responses as a way of leading her mathematics discussion. She said that it would have been enough for her to ask one student to show his or her work on the board and explain the answer. However, the question was framed in a way that signaled her to recognize that multiple students' perspectives were valued for use in mathematics discussion.

Third, the interviewees pointed out the mismatch between the mathematical problems used in the items and the grade level the mathematical problems were assigned to. The mathematical problems used in item 5, item 6, and item 10 were reported to be inappropriate because of the perceived difficulty of the problems for second graders. As a result, it was difficult for the interviewees to think about how the teacher should lead a discussion in such scenarios, and they chose more direct instruction in which the teacher explains and shows the students how to arrive at the correct answer.

Fourth, the lack of visual aids and the unclear use of visual aids in the BMD measure were also reported to be problematic. Three interviewees shared the view that, for items 5 and 6 in which there are three students' mathematical work, it would have been helpful, whether correct or incorrect, to have been provided with the work itself instead of having to imagine it. They also said that drawing coins to represent students' answers took too much time before they could consider how the teacher should lead a discussion. Two out of three interviewees pointed out that seeing students' work was critical for them to determine how the discussion should proceed, but the item 5 and 6 did not have students' work available. Similarly, the visual aid of base ten blocks also caused some confusion. The graphic representation of base ten blocks was originally presented to help respondents who were not familiar with what base ten blocks were. However, one interviewee did not understand the purpose of the graphic representation of base ten blocks, since it was not the representation of the number 406 in the problem.

Lastly, the response options for some items did not match the problem statement. For example, item 8 was problematic for several reasons according to all interviewees. First, the problem was not clear regarding the multiples of which number the students were asked to find, yet the third response option included information about multiples of 4 and multiples of 8. Another interviewee pointed out that the first response option (ask for answers then highlight correct answers while reviewing the definition of multiple) included too many activities. For item 10, one interviewee said that the response options were too similar to one another. The findings from this section can be used to revise the instrument for improvement in another iteration of instrument development.

Research Question 3 (RQ 3): What is the Evidence for Internal Structure Validity?

Internal structure validity refers to the degree to which the statistical structure of items and respondent is consistent with the definition of the construct and intended logical structure of the instrument (APA, AERA, & NCME, 1999). The Wright Map and item fit statistics from the

ConQuest (1998) output were used to analyze the fit of the items and determine if the empirical results of the Wright Map (Wright & Masters, 1982) support the theoretical framework in the construct map (Wilson, 2005).

Wright Map.

The Wright Map (Wright & Masters, 1982) is a tool used to produce a visual representation of data (Wilson, 2004). It is an empirically calibrated version of the construct map used to gather evidence based on internal structure of the construct that is hypothesized (RQ 3: what is the evidence for internal structure validity?). The Wright Map allows us to check whether the empirical results support the theoretical expectations in the construct map. The Wright Map in Figure 6 was used to check the consistency and distinction of four levels.

It was expected to find four levels in the Wright Map as the theory suggested, but the distinction between the level 3 (emerging dialogical) and level 4 (dialogical) was not as clear as it should have been. In the Wright Map in Figure 9, the estimated respondent parameters are on the left side and the estimated item parameters are on the right side after calibrating the items using the Rasch Model. The Xs on the left hand side of the map represent the proficiency of 168 respondents as distributed across the sample. There are ten items represented on the right side, with their respective levels from the construct map. Each X represents 2 cases. Overall, the distinction between four levels of teachers is reflected in the Wright Map.

If a respondent (e.g., 0 logit) has the same location as a particular item (e.g., 7.2) on the Wright Map, it means that this respondent has almost 50% chance of expressing a belief that represents univocal or partial univocal belief about mathematical discourse. However, if a respondent's location is above the item (e.g., 1 logit) on the Wright Map, this respondent has a greater than 50% chance of expressing univocal or partial univocal belief about mathematical discourse. On the other hand, if a respondent is located below the item (e.g., -1 logit) on the Wright Map, this respondent has a less than 50% chance of expressing a univocal or partial univocal belief about mathematical discourse.

The Wright Map shows a distribution of person estimates that is approximately unimoded. The range of person estimates extends from less than -3 to 3 logits, which suggests that the instrument succeeds in placing individuals along a reasonably sufficient continuum. The distribution of item difficulties covers the same region as the distribution of student proficiencies. That is, there are respondents represented at every level where there are items to measure them. However, the range of estimates for individuals extends slightly wider at the upper level than that of the item thresholds. Looking at respondent location, most teachers were unlikely to express univocal belief about mathematics discussion for most items. The four distinct groups of item thresholds indicate a level of internal construct validity.

However, level 3 (emerging dialogical) and level 4 (dialogical) for most items overlap, and the distinction between the two is not as clear as for level 1 and level 2. The overlapping of level 3 (emerging dialogical) and level 4 (dialogical) was not, however, surprising because the description for level 3 (emerging dialogical) and level 4 (dialogical) was not substantially different. For example, teachers at level 3 (emerging dialogical) were characterized as encouraging students to share their mathematical reasoning with one another, whereas level 4 teachers (dialogical), the highest level of the construct map, was characterized as valuing the opportunity of students sharing different ways of solving the problems, but also took it one step further by encouraging students to communicate their mathematical ideas with one another, through debating and agreeing or disagreeing.

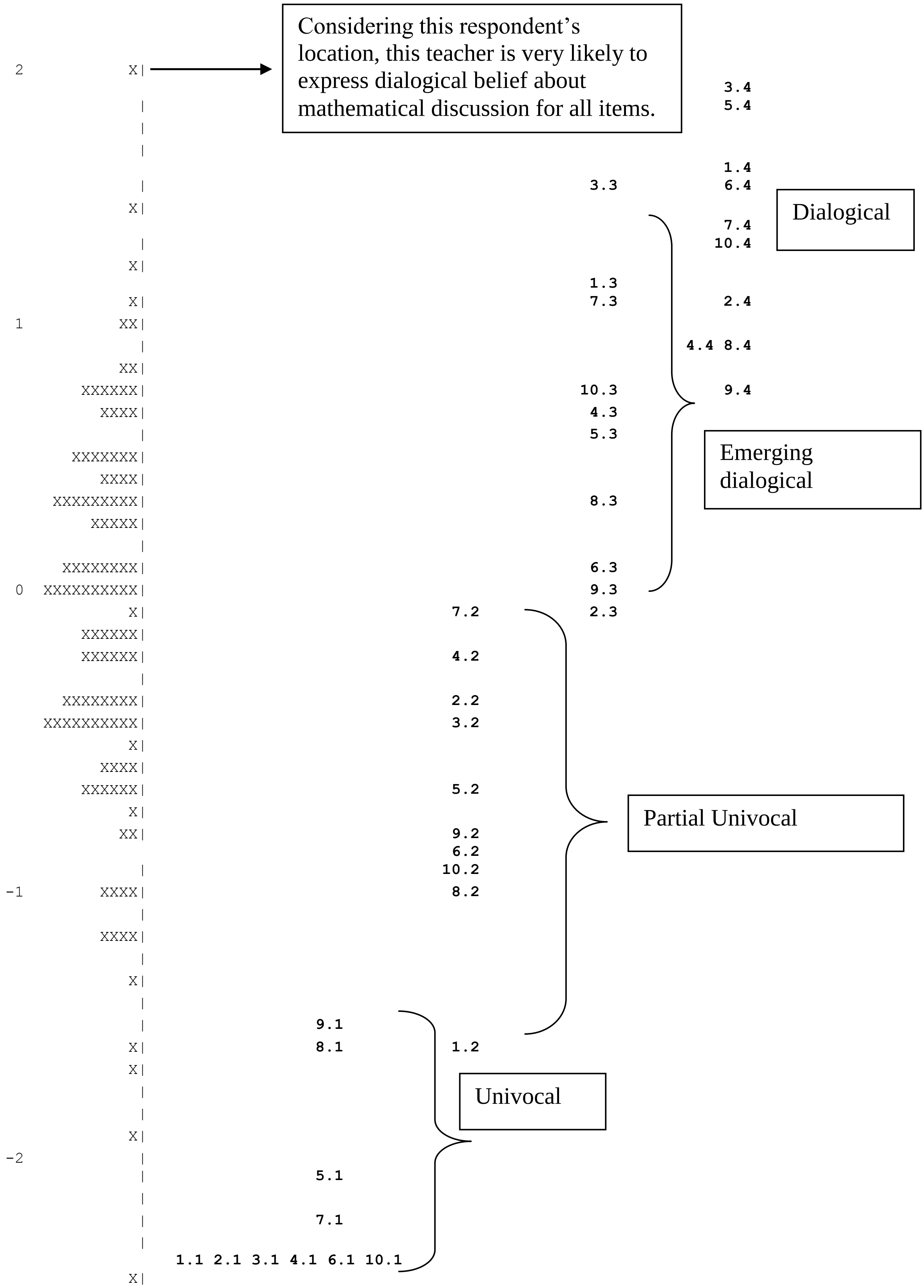


Figure 6. *Wright Map organized by thresholds*

For most items, the distance between level 3 (emerging dialogical) and level 4 (dialogical) was small, whereas the distance between level 1 (univocal) and level 2 (partial univocal) was larger. Appendix I helps to illustrate this point; it shows the Wright Map as organized by all the items. Figure 7 below only shows the Wright Map of the three items (#1, 2, and 9) to support the following discussion. I will present a few typical responses that are scored as each level.

<table>
<tr><th>Distribution of students</th><th colspan="3">Distribution of Item</th></tr>
<tr><th></th><th>Item 1</th><th>Item 2</th><th>Item 9</th></tr>
<tr><td>XX|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td></td></tr>
<tr><td>2 |</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td></td></tr>
<tr><td>X|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td>4</td><td></td><td></td></tr>
<tr><td>X|</td><td></td><td></td><td></td></tr>
<tr><td>X|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td>4</td><td></td></tr>
<tr><td>XX|</td><td>3</td><td></td><td></td></tr>
<tr><td>1 XX|</td><td></td><td></td><td></td></tr>
<tr><td>XX|</td><td></td><td></td><td></td></tr>
<tr><td>XX|</td><td></td><td></td><td></td></tr>
<tr><td>XXXX|</td><td></td><td></td><td></td></tr>
<tr><td>XXX|</td><td></td><td></td><td></td></tr>
<tr><td>XXXXX|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td>4</td></tr>
<tr><td>XXXXX|</td><td></td><td></td><td></td></tr>
<tr><td>XXXXXXXXXX|</td><td></td><td></td><td></td></tr>
<tr><td>XXXX|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td></td></tr>
<tr><td>XXXXXX|</td><td></td><td></td><td></td></tr>
<tr><td>0 XXXXXXXX|</td><td></td><td></td><td>3</td></tr>
<tr><td>XXX|</td><td></td><td>3</td><td></td></tr>
<tr><td>XXXX|</td><td></td><td></td><td></td></tr>
<tr><td>XXXXX|</td><td></td><td></td><td></td></tr>
<tr><td>X|</td><td></td><td></td><td></td></tr>
<tr><td>XXXXXX|</td><td></td><td>2</td><td></td></tr>
<tr><td>XXXXXXXXX|</td><td></td><td></td><td></td></tr>
<tr><td>XX|</td><td></td><td></td><td></td></tr>
<tr><td>XXX|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td></td></tr>
<tr><td>XXXXX|</td><td></td><td></td><td></td></tr>
<tr><td>X|</td><td></td><td></td><td></td></tr>
<tr><td>X|</td><td></td><td></td><td>2</td></tr>
<tr><td>-1 XXX|</td><td></td><td></td><td></td></tr>
<tr><td>X|</td><td></td><td></td><td></td></tr>
<tr><td>XXX|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td></td></tr>
<tr><td>X|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td>1</td></tr>
<tr><td>X|</td><td>2</td><td></td><td></td></tr>
<tr><td>X|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td></td></tr>
<tr><td>X|</td><td></td><td></td><td></td></tr>
<tr><td>-2 |</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td></td></tr>
<tr><td>|</td><td></td><td></td><td></td></tr>
<tr><td>X|</td><td>1</td><td>1</td><td></td></tr>
</table>

Figure 7. Wright Map organized by item

A kindergarten teacher asks students to copy a pattern series and fill in the blank to continue the pattern.

<u>Problem on the board:</u>

△ □ ○ △ □ ○ △ □ ○ ___

The teacher reconvenes the class and asks Sarah to come up to the board to fill in the blank. Sarah fills the blank with a circle.

What should the teacher SAY and DO to guide students toward an understanding of the correct pattern?

Figure 8. Item 1

Teachers at level 1 (univocal) typically acknowledge Sarah's contribution, yet immediately tell Sarah or the class that the answer is incorrect. Teachers at this level explain or show why a circle is not the correct answer. Following are three types of typical responses of level 1 teachers:

> "*Often times when students explain something it can be a bit confusing so the teacher should try to repeat the explanation in their own words*, pointing out the important parts of the number sentence. Such as when it says 'how much does she have left'…the teacher should explain that this type of wording often means it's a subtraction equation."
>
> "Teacher should use a tens and ones chart along with manipulatives (beans, units, etc.) to show that you cannot subtract 8 beans from 2 beans. Then *teacher should show with many examples over and over again*, how to regroup tens into ones to subtract. Teacher should identify the incorrect ones by stating what is incorrect and what the student might do to correct their error."
>
> "*The teacher can start by explaining* that multiplication is the addition of groups. Then *show* the groups and then add them. Then *show* them that multiplication is a shorter to add groups. Then *explain* the multiplication process."

As illustrated in Figure 8, in item 1, in order for the teacher to value multiple students' answers and solutions (level 3: emerging dialogical) in a discussion, the respondent location has to be at least 1.0 logits. Only 9 respondents were at the respondent location of 1.0 logits or above out of 167 respondents.

On the other hand, teachers at level 2 (partial univocal) focus more on providing the student (typically the one who presented their answer on the board) with the opportunity to explain his or her answer and method instead of providing the student with the correct answer. Teachers at this level tend to assist the student to do this by asking a series of questions for the student to answer. This is very much like the series of "steps that comprise a solution path" (Lloyd, 2005) or "the teacher-student interaction followed an intiate-respond-evaluate (IRE) pattern focused on eliciting final answers" (Cross, 2009). Teachers at this level tend to focus on a single student rather than the whole class, and thus the conversation is more between the teacher and one particular student. Typical responses of level 2 teachers include the following:

> “The teacher should ask Ben [the student] for his attempt and ask him to have a seat. I would review how to complete the subtraction problem with the students, going over it step by step.”

> “[The teacher should] ask questions about the properties of triangles to check for student understanding. This way the students will inform the teacher of their misunderstanding so they [the teacher] can specifically correct it.”

> “Either have Leandra or the teacher write out the number sentence on the board and ask the class what clues in the sentences told Leandra to write the equation that way. Get to explaining that “how much” and “have left” are key clues.”

Teachers at levels 2 and 3 express very specific ideas about what students should know as a result of their classroom work (Lloyd, 2005). Teachers at level 2 frequently expect that having the student explain their own answer would lead them catch their own mistake. Level 2 teachers clearly express their belief that giving out the correct answer to the student is not desirable, but they expressed that Sarah’s incorrect answer was due to their lack of paying close attention to the problem.

Teachers at level 3, on the other hand, express the importance of helping the student(s) who got the incorrect answer, like Sarah in item 1, by creating the opportunity for multiple students to present and share their answers and reasoning. Level 3 teachers express the importance of correcting the student’s answer through a series of questions. They also emphasize their interest in assessing multiple students’ mathematical thinking. The approaches used by teachers at this level are less guided and structured by their specific questions. Level 3 teachers state the importance of students sharing their answers with one another, and listening to multiple answers and methods from one another. Typical responses of level 3 teachers following include:

> “Show [the class] how each person can come with different ways to get the same answer. Ask each student to show why they chose their ways of adding up to a dollar.”

> “Ask students if they were able to come up with the answer in a different way. Check to make sure that all students are on the same page before continuing.”

> “By allowing all the students to present their ideas the teacher has an understanding of what the students know. The students can also hear other students thoughts and therefore either change their thoughts or gain a better undersatnding of the material. Students will have a chance to explore the problem and come up with solutions rather than always relying on the teacher for explanations. After exploring the problem the class can come up with the multiples and how many they can find as a way to find the final solutions.”

Teachers at this level value students’ participation in learning mathematics because they see this as a way to transform the social norm about what it means to learn mathematics.

Teachers at level 4 express the importance of students’ ability to articulate their mathematical answers and solutions, not only to the teacher, but to their peers, and emphasize the creation of an environment in which students play an active role in teaching and learning from one another. Level 4 teachers encourage students to explain, defend, and challenge their answers and methods of solution through agreement and disagreement with one another’s ideas. Typical responses of level 4 teachers are following:

> "It is important to get responses from several students, but it is also important to have the rest of the class agree of disagree with their classmates. This gives students the opportunity to hear multiple viewpoints, teach their peers, and actively particpate in the class discussion."
>
> "Ask if each combination equals one dollar. Have students vote on which pair came up with the best answer according to the problem. Have a student representative for each group explain why they thnk it is the best and then have the class re-vote."
>
> "I think the teacher should ask Sarah to explain why she chose the circle. If Sarah does not give a response, the teacher should let her sit down and ask another student in the class why they think Sarah chose that option. The teacher can then ask and confirm Sarah's thoughts from the other students. Then the teacher can ask other students if they agree with Sarah's prediction. The students that disagree, and chose a triangle, should be called on and give their reasoning."

Teachers at level 4 express the importance of students taking charge of their learning, and encourage them to not simply settle for the final answer but to examine their own thinking and the thinking of their peers (Cross, 2009).

As illustrated in the Wright Map (Figure 7), moving from level 1 (univocal) to level 2 (partial univocal) was relatively easy, but moving from level 2 (partial univocal) to level 3 (emerging dialogical) was disproportionately more difficult. Theoretically, moving from level 2 (partial univocal) to level 3 (emerging dialogical) is a significant shift for teachers to make from attending to a single student to multiple students in leading a discussion. Teachers may believe that they are attending to multiple students just because they are teaching a whole class, but in level 2, the conversation takes place often between the teacher and a particular student, one at a time, with little attention to get the rest of the class involved. Moreover, it is significant shift for teachers because the purpose of conversation is not limited to guiding the students to arrive at the correct answer, but encouraging them to explore and present different answers and methods, as well as supporting students in listening to each other. For many teachers who believe mathematics is a set of operations and the primary goal of instrument is to correct incorrect answers, level 3 describes a new norm for learning mathematics.

In item 1 (Figure 8), many teachers expressed in their open-ended response that kindergarten students were too young to engage in mathematics discussions. They expressed that direct instruction through the step-by-step procedure with Sarah would be the most effective way to lead a discussion because of Sarah's grade. More than 80% of the teachers responded that they would either show how to find the correct answer or assist the students to arrive at the correct answer through a step-by-step process.

In contrast, for item 9 (see Figure 9), it was much easier for teachers to express their dialogical approach to mathematical discourse because of the mathematical problem used in item 9. Teachers expressed that sorting polygons into groups was more amenable to mathematical discussion than correcting students' incorrect answers with a subtraction problem in item 3 and a multiplication problem in item 4. Thus teacher belief about mathematical discourse may interact with the affordances of the mathematical problem for discussion.

A third grade teacher launches a geometry unit with the following sorting task.

[Problem] Sort your polygons into two groups: triangles and other shapes. Students work in pairs. Their polygons include triangles (scalene, equilateral, isosceles, and right triangles) and other polygons (squares, rectangles, hexagons, and parallelograms). The teacher notices that some students sort into three groups: equilateral triangles, all other triangles, and
all other polygons.

In this situation, the MOST important thing the teacher should do with the class is…..

Explain why your choice is most important.

Figure 9 *Item 9*

Item fit analysis.

Table 10 shows the fit of the items, including the calibrated item parameters, the infit mean square, and the t-values. The mean square fit statistic (Wright & Masters, 1981) allows us to detect differences in the actual residuals compared to a theoretical expectation of their variance, assuming the empirical data fit the Rasch item response model well. A generally accepted effect size (Adams & Khoo, 1996) for the weighted mean square value would have a lower bound of .75 (=3/4) and an upper bound of 1.33 (=4/3). According to Table 4, all infit mean squares fall between 0.75 and 1.33. Infit mean square values less than 0.75 indicate that items exhibit more predictable responses than expected while mean square values greater than 1.33 indicate that items exhibit more random responses than expected. Based on the infit mean square, and the t-values, all ten items fit acceptably well. Figure 9 is a visual representation of the weighted infit mean square for each item. All infit mean squares fall between 0.75 and 1.33, as indicated in the yellow shaded area.

Item	Infit Mean Square	t-Value	Good Fit?
1	.98	-0.1	Yes
2	.99	-0.1	Yes
3	.94	-0.4	Yes
4	1.00	0.0	Yes
5	.94	-0.6	Yes
6	1.00	0.1	Yes
7	0.98	-0.2	Yes
8	1.10	1.0	Yes
9	1.09	0.9	Yes
10	1.01	0.1	Yes

Table 10 *Item Calibration Estimates and Fit Statistics*

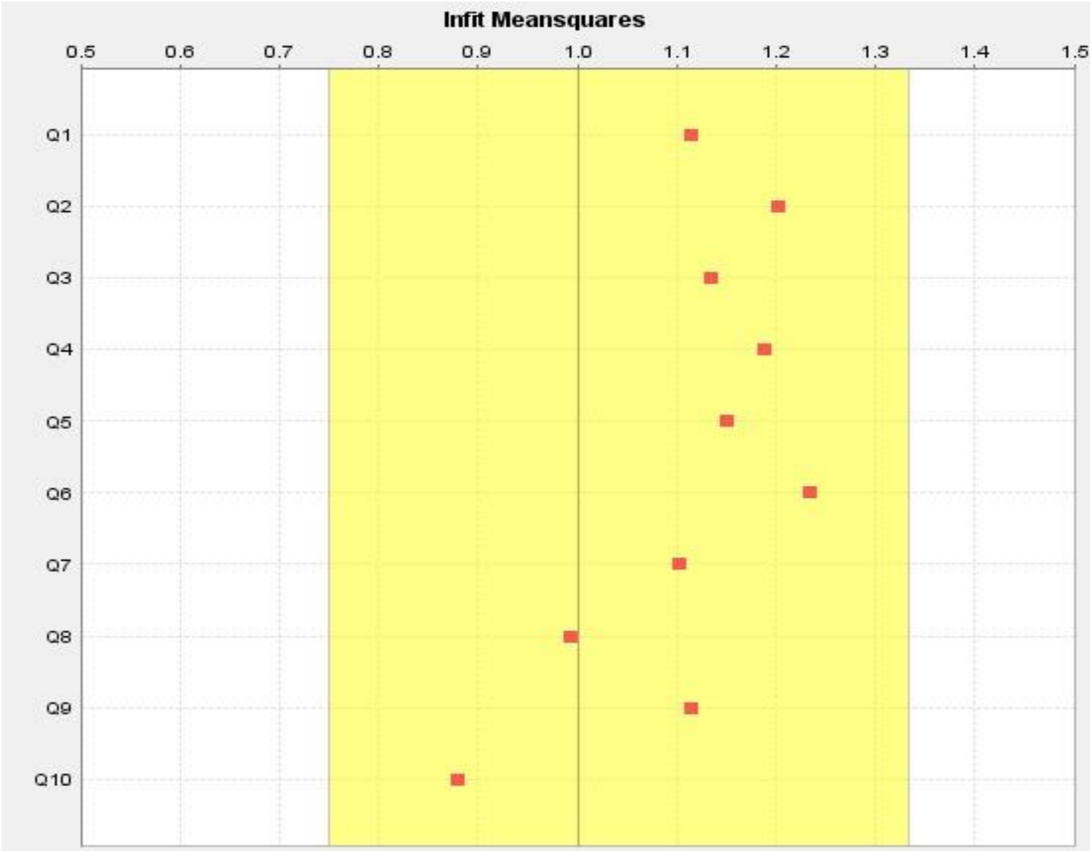


Figure 9 *Item Number as a Function of Weighted Infit Mean Square*

The *weighted t-statistic* uses a transformation that attempts to make the weighted mean square into a standard normal distribution. The weighted t-statistic is sometimes used to test the statistical significance of the mean square (Wright & Masters, 1981). All t-statistics fall between -2 and 2. Any t-statistic less than -2 or greater than 2 indicates a statistically significant misfit. Thus according to the t-statistics, the BMD items do not show any signs of a statistically significant misfit.

The results of the item analysis show that individual items were generally consistent with the BMD measure as a whole. Appendix K presents information on the mean person locations and the standard error of those locations generated by the partial-credit Rasch IRT model. Appendix K also provides the results for traditional statistics that include: (1) the number of response categories for each item, (2) the number (count) of respondents who answered in each of the categories, (3) the percentage of respondents who answered in each of the categories, and (4) the point bi-serial correlation for each response category. In general, it is expected that respondents who score at a higher category on an item will also score higher on the BMD (Wilson, 2005). An example is shown for item 1 in Table 11. The mean locations of all ten items are increasing in the expected way (Appendix K).

Item 1
------
item:1 (1)
Cases for this item 168 Discrimination 0.49
Item Threshold(s): -3.82 -1.61 1.11 1.45 Weighted MNSQ 0.98
Item Delta(s): -3.70 -1.70 2.28 0.24

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	1	0.60	-0.11	-1.40(.164)	-1.03	0.00
1	1.00	25	14.88	-0.25	-3.36(.001)	-0.41	0.60
2	2.00	113	67.26	-0.16	-2.05(.042)	-0.08	0.63
3	3.00	13	7.74	0.23	3.03(.003)	0.53	0.38
4	4.00	16	9.52	0.38	5.23(.000)	0.91	1.04

Table 11 *Item Analysis*

Research Question 4 (RQ 4): What is the Evidence for External Validity?

According to the *Standards for Educational and Psychological Testing* (APA, AERA, NCME, 1999), “analyses of the relationship of test scores to variables external to the test provide another important source of validity evidence” (p. 13). To address the fourth research question (RQ4) about validity evidence based on relations to variables external to the BMD measure, the study examined the convergent evidence between the data gathered using the BMD measure and the data gathered using the Preservice Teachers’ Attitudes about Discourse in the Mathematics Classroom (PADM) measure developed by Casa, McGivney-Burelle, and DeFranco (2008). Since the content of the PADM instrument and the BMD measure appeared to measure a similar construct about teachers’ attitudes and beliefs about mathematics discourse, it was hypothesized that the statistics would show a reaonably high correlation between the two measures. The reliability of the PADM measure is 0.89 using the sample of this study. The analysis of each item on the PADM measure is included in Appendix L. Descriptive analysis of the 25 Likert-scale questionnaire items in the PADM measure is shown in Appendix O. The reliability of the PADM measure indicates an adequate level of internal consistency of the measure.

There are two procedures used to calculate the correlation between two measures to demonstrate evidence of external validity: staged procedure and direct procedure. The staged procedure for correlation is a two step consecutive procedure. In the first step, the two measures, PADM and BMD, are calibrated consecutively to yield two respondent locations and two estimated error variances for each respondent. In the second step, the Pearson correlation between the two respondent locations and reliability are calculated. When two sets of measures are correlated, however, measurement error lowers the correlation coefficient below the level it would have reached had the measures been precise. Since the reliability of a set of measures is the proportion of observed variance not due to measurement error, (Schumacker & Muchinsky, 1996), a disattenuation formula can be applied to correct the correlation. Disattenuation of correlation allows us to estimate how large the correlation would have been if the latent traits had been measured without error. Thus from a correlation coefficient, *rPADM & BMD* , measurement error can be removed to estimate the correlation coefficient disattenuated of measurement error, *RPADM & BMD*, by the formula (Spearman 1904, 1910):

$$RPADM \ \& \ BMD = \frac{r\text{PADM} \ \& \ BMD}{\sqrt{r\text{PADM} \ r\text{BMD}}} \tag{4}$$

On the other hand, improved parameter estimates and individual measurements are possible by direct procedure for correlation, using the Multidimensional Random Coefficients Multinomial Logit model (MRCML), which draws on the relationship between the latent dimensions (Adams, Wilson, & Wang, 1997). The results of these two procedures for calculating correlation are compared and presented in Table 12.

Model	Reliability
Consecutive (staged procedure)	0.00
Multidimensional	0.01

Table 12 *Reliability Indices*

The expected correlation between the two measures was not found. On the contrary, the correlation between the two, indicated in Table 12, shows that the two measures are measuring two orthogonal constructs. The reliability calculated using the multidimensional model was slightly higher, although negligibly, than that of the consecutive model, because direct estimation yields unbiased parameter recovery. The multidimensional procedure is more efficient than the consecutive model in which two measures are analyzed separately for each dimension and yet fail to use all available data (Wang, 1999). The Wright Map of the multidimensional IRT analysis is presented in Appendix J. Although the items on the PADM instrument measure a wider range of the respondents in their beliefs than the BMD measure, most PADM items were easy to agree with for the respondents. This finding is also confirmed by the distributions of the responses of the PADM measure that a majority of the respondents strongly agreed or agreed with the statements that support the importance of mathematical discussion in elementary classrooms (Appendix O).

To investigate the lack of correlation between the two measures, two analyses were performed. The first analysis examines how teachers responded to the PADM measure and BMD measure to identify any consistent, inconsistent, or contradictory evidence. Thus post-hoc think-aloud interviews were conducted with five of the study participants. Although, according to the PADM measure, the respondents expressed their belief that teachers should encourage mathematics discussion in the mathematics classroom, the extent to which teachers encouraged students to debate mathematical ideas during instruction and the extent to which teachers encouraged students to take a lead in mathematics discussion varied depending on the particular teaching scenarios presented in the BMD measure. For example, all five study participants expressed that students in lower grade levels can share their answers with one another, but they expresed their concern that discussion would be challenging for students in lower grade levels because of difficulty in articulating mathematical concepts. Similarly, they also expressed that mathematical discussion would be more appropriate and effective for students in reviewing materials taught previously than when students are introduced to a new concept for the first time.

Moreover, some participants' expressed actions did not necessarily match with the general belief of what a mathematics discussion should look like. This was found when the respondents were probed further about how and when the teacher should lead a mathematics discussion using the BMD measure. For example, one respondent said that he "strongly agrees" with the statement in the PADM measure, "students should exchange their mathematical ideas with other students." However, when he answered item 4 in the BMD measure[6], he wrote, "*Often*

[6] The multiplication problem (7X15) solutions presented by Niral, Carlton, and Jasmine (See Appendix G).

times when students explain something it can be a bit confusing so the teacher should try to repeat the explanation in their own words." When this teacher was further probed during a post-hoc think-aloud interview, he elaborated that students exchange their mathematical ideas with each other, but the teacher should explain the correct method and answer to students because student-led discussion lacks the mathematical precision (Nathan & Knuth, 2003). Using specific teaching scenarios in the BMD measure demonstrated that teachers' beliefs about mathematical discussion are more complex than simply agreeing or disagreeing with a series of statement about what they believe about mathematical discourse.

These examples suggest much more than just that the BMD measure shows that teachers' beliefs are complex. They also suggest that the PADM measure may be much more subject to social desirability bias than the BMD measure. Social desirability bias is a systematic error in self-report measures results from the desire of respondents to avoid embarasment and project a favorable image to others (Fisher, 1993). Social desirability bias is pervasive in social science research because of the basic human tendency to present oneself in the best possible light (Fisher, 1993). To mitigate the effects of social desirability bias, indirect questioning has been used frequently in social sciences. Indirect questioning is a projective technique that asks respondents to answer structured questions from the perspective of another person, and thought to reduce the distortion of private opinions by asking respondents to report on the nature of the external world rather than about themselves (Anderson 1978; Calder and Burnkrant, 1977; Westfall et al. 1957). For this reason, the BMD measure consist of hypothetical teaching situations that present students' correct and incorrect thinking, and to then ask the teacher to respond in order to lead a mathematical discussion for the purpose of compensating for social desirability bias. Although there are good theoretical reasons to believe that indirect questioning reduces social desirability bias, empirical research on the subject is limited and inconsluvie (Fisher, 1993).

To examine if teachers answered to the items in a way they believe is socially acceptable and desirable, the average scores of the BMD measure and the PADM measure using 163 respondents[7] were calculated and compared to examine the difference between the two scores. If social desirability bias were to be an issue, it would be the case that teachers' positive beliefs about mathematical discourse would be more inflated when their beliefs are assessed using the PADM measure. As expected, out of a total of 163, only 4% (6) of participants showed slightly higher scores of the BMD measure than the PADM measure although the differences were negligible, ranging from -0.1 to -0.4. On the other hand, 96% (157) of participants showed higher scores of the PADM measure than the BMD measure, and the differences were wider, ranging from 0.1 to 2.9. This suggests that teachers appear to be much more supportive of students engaging in mathematical discussion in elementary classrooms when their beliefs were assessed using the PADM measure than the BMD measure. Appendix P presents information on the mean scores of the BMD measure, the mean scores of the PADM measure, and the difference between the two measures.

The second analysis includes the examination of a selected 12 items out of total 25 that are the most closely related to the BMD measure based on the content of the items (Appendix N). Thereafter, multidimensional analysis is perfomed. The following is the list of selected items from the PADM measures that are mostly related to mathematical discussion. Correlation using direct procedure was 0.08, which showed a slight improvement yet still neglible.

[7] Out of a total 167 participants who completed both the BMD measure and PADM measure, 4 cases were dropped due to too many missing responses to the items.

Research Question 5 (RQ 5): What is the Evidence for Reliability (Internal Consistency)?

Reliability, which is an integral part of validity, indicates the degree to which the instrument measures the intended construct with sufficient consistency over individuals for intended usage (Wilson, 2005). It is important for the instrument to function with sufficient consistency over respondents. The value of reliability ranges from 0.0 to 1.0. As the reliability gets closer to 1.0, it suggests that the survey instrument performs consistently well across respondents. Three reliability indicators are examined in this study: (a) Cronbach's alpha (Cronbach, 1990), and person separation (Wright & Masters, 1982), (b) measurement error, and (c) inter-rater reliability. These indicators suggest that the reliability of the BMD measure is high.

Cronbach's alpha.

In classical test theory, Cronbach's alpha (1951) is the standard reliability index for internal consistency coefficients. It is calculated from the correlations between pairs of items. Cronbach's alpha is a coefficient, ranging between 0 and 1, which is used to rate the internal consistency or the correlation of the items in a test. In Rasch IRT modeling, the equivalent indicator for Cronbach's alpha is person separation reliability (Wright & Masters, 1981), which describes how well the variability in the data from a single administration of the test can be explained by the measurement model. A statistic closest to '1' represents a strong proportion of variance that can be accounted for by the model. The output of the ConQuest software provides both types of internal consistency coefficients. Both reliability indices indicated a relatively high consistency of the BMD measure.

Coefficient Values (r)	
Cronbach's Alpha	Person Separation Reliability
0.72	0.75

Table 13 *Internal Consistency Reliability Coefficients*

Measurement error.

The standard error of measurement (SEM), the standard deviation of the estimated ability θ, is a very useful tool for assessing the accuracy of an estimate of a respondent's location. The respondents' score may differ from their true score because of different influences on the score, which may be due to various conditions (e.g., testing condition, room temperature, fatigue, the length of the instrument). ConQuest provides SEM for all parameters. The SEM indicates how well a respondent is measured depending on her or his location to an item. Thus in IRT, every estimated score has its own SEM, because it is assumed to vary with the ability level of respondents. As an instrument's standard error decreases, its reliability coefficient increases. As an instrument's standard error of measurement increases, the instrument becomes less reliable. Figure 10 shows the SEM for the BMD measure.

In Figure 10, the SEM for the respondents shows the distribution of the items over the construct. SEM is consistent with the respondents' narrow distribution, indicating higher levels of error at the extreme end of the scale where there is generally less respondent information. Thus the SEM (*SEM(θ))* is smaller in the middle than at the end.

The smallest SEM is about 0.25 logits. Locations for respondents within two logits of zero are estimated with a SEM between .25 and .4. However, the SEM curve is discontinued between around .5 logit to 1.0 logit. This means that the BMD measure is not useful to measure respondents whose belief range falls between .5 logit to 1.0 logit.

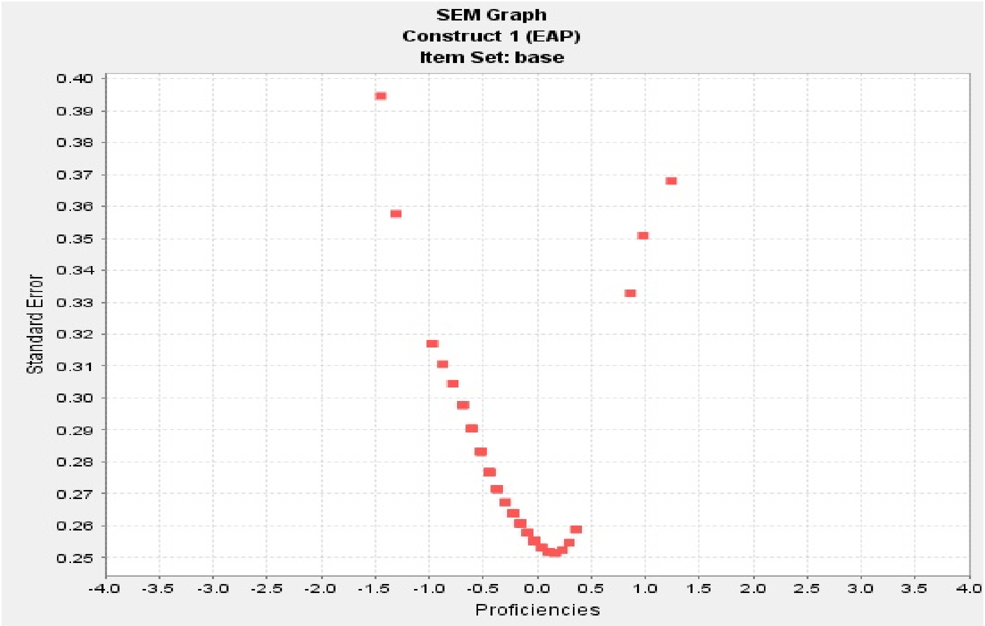


Figure 10. Standard error of measurement for the BMD measure

The other way to express the relationship is the test information function. The information (*Inf(θ)*), the reciprocal of the square of the *SEM*, expresses this relationship (Lord, 1980). Figure 11 shows the information (*Inf(θ)*): $(Inf(\theta)) = 1/SEM(\theta)^2$

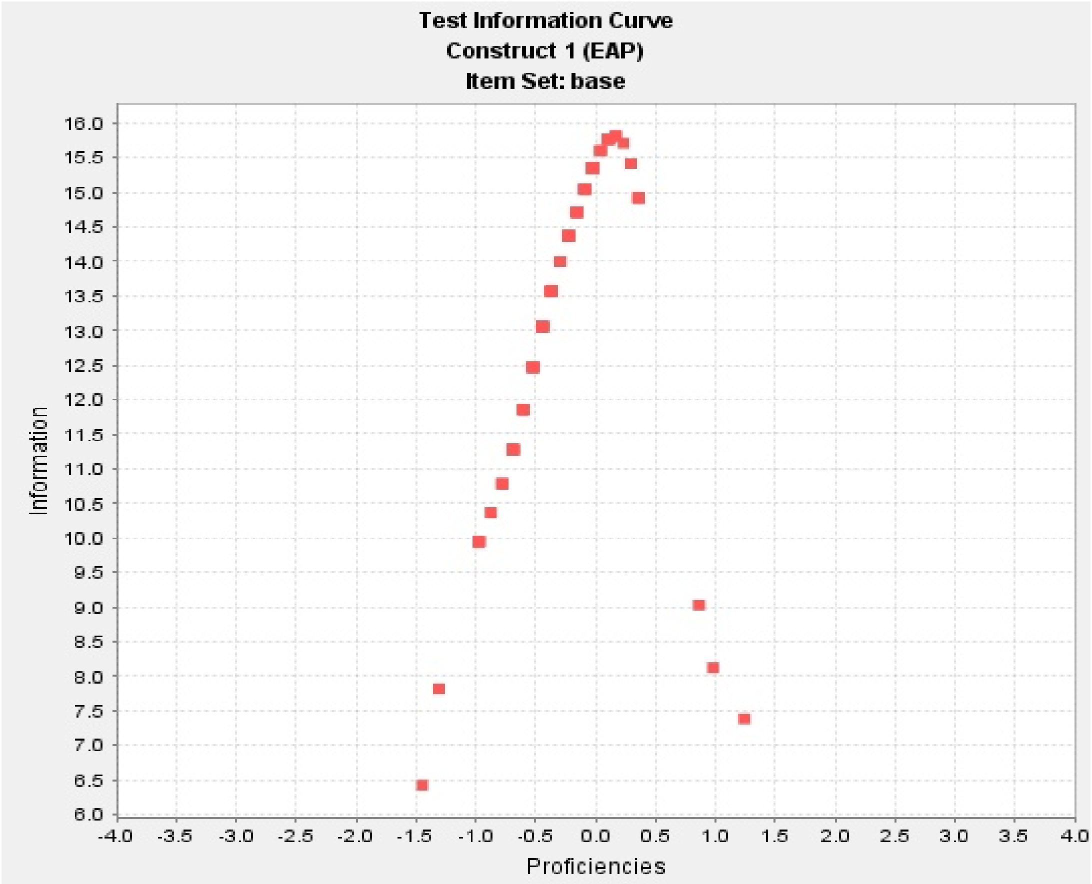


Figure 11. The information for the BMD measure

The information *(Inf(θ))* is less interpretable than the standard error of measurement curve. According to Figure 11, the most robust part of the BMD measure is from approximately -1.5 to 1.5 logits. The distribution of the respondents in Figure 8 shows that many respondents in this sample are between -1.5 to 1.5 logits. The minimum information is 0.5 and the maximum information is 9.0 based on the set of 10 items. The shape and range of the information curve suggest that the BMD measure functions optimally for a large proportion (96%) of the sample. However, when the respondent location was higher than 1.5 logits, the BMD measure did not function well.

Inter-rater reliability.

Inter-rater reliability is the degree to which raters agree in the scoring of the items. The investigation into rater consistency is important since the BMD instrument collects data using open-ended items. A total of 3 scoring sessions were conducted in the month of Feburary 2010. Prior to each scoring session, steps were taken to reduce rater inconsistency with the inclusion of a rater training protocol (Wilson, 2005). The training protocol for each session included: (a) background information on the item tasks and their relation to the construct map for a given dimension, (b) a preliminary moderation session for the raters to examine, review, and score a wide range of responses, (c) examination of item response exemplars that could be clearly assigned to a category or level of response on a scoring guide, and ones that were not so clear, and (d) discussion of ratings on specific item responses. The relationship between the two sets of ratings is examined. Ten percent of the data (or 20 surveys) was double-scored by two different researchers to determine the extent of agreement. Two sets of independent scores for the open ended items were derived from two raters in this study.The number of items double scored and the correlation between the rater scores are presented in Table 14.

Name of Item	Mathematical Content Used in the Items	Correlation Between Scores of Rater 1 and Rater 2
1	Pattern	0.97**
2	Subtraction	0.60**
3	Subtraction	0.90**
4	Multiplications and place values	0.97**
5	Additions	0.84**
6	Using base 10 blocks to represent the whole number	0.83**
7	Subtraction and place values	0.80**
8	Concept of multiple	0.50*
9	Sorting geometric figures based on their property	0.69**
10	finding solutions for pairs of equations that are number sentences	0.89**
** statistically significant at the .01 level * statistically significant at the .05 level		

Table 14 *Correlation Between Scores for Rater 1 and Rater 2 on Open-Ended Items*

Six of the ten items show statistically significant correlations between the scores of the two raters, with ten of the items indicating correlations above 0.75, demonstrating a reasonable level of consistency in scoring across researchers for those items. Rater correlations for two items, “subtraction” and “the concept of multiple,” were relatively low. The situation in which the first student the teacher called on, Leandra, offered a correct answer to the problem, and the respondent was asked to share how the teacher should respond to lead a discussion in the situation. 2 raters discussed the discrepancy between their scores for item 2. It was found out that Rater 2 (the one who received the rater training) had a tendency to rely on some keywords or phrases to determine the score. Consider, for example, the following response to Item 2:

> “Thank you, Leandra. Does anyone have a question for Leandra about her answer or how she got it?” If not, move on and ask the students to do the problem at their desks and see if anyone got a different answer and/or explanation.

Rater 2 scored this response as 1, because the teacher was willing to move on instead of showing more interest in leading a discussion using or based on Leandra’s answer; the word “move on” was the key signal to score this response 1. Rater 1 scored this response as 4, because the teacher facilitated the discussion among students by encouraging the class to ask questions to Leandra about her answer.

Another source of low correlation was the lack of clear descriptions to differentiate adjunct levels. When the response included raters often disagreed. To illustrate how two raters disagree with the response, one respondent wrote, “[The teacher should] ask if anyone got a different response and then lead a discussion about why each student came to each conclusion”. Rater 1 scored this as 3 because the teacher explored differences in students’ thinking; level 4 was rejected, because the response did not elaborate on what discussion could look like. In contrast, Rater 2 scored this as 4 because of the teacher’s expressed interest in students having a discussion about how they came up with their answers.

A *level of agreement* was calculated across all items, by calculating the difference between the score from Rater 1 and the score from Rater 2. If the level of agreement was '0' there was complete agreement in the scores. If the level of agreement was '1' then Rater 1 scored the item response one level higher than Rater 2. As indicated in Figure 10, the level of agreement ranges from -3 to 4, but the difference between the scores of Rater 1 and Rater 2 was '0' across 149 out of 168 items scored. This represents an exact agreement in scores for 89% of the data. This provides strong evidence that the scoring guides were being utilized in a consistent way.

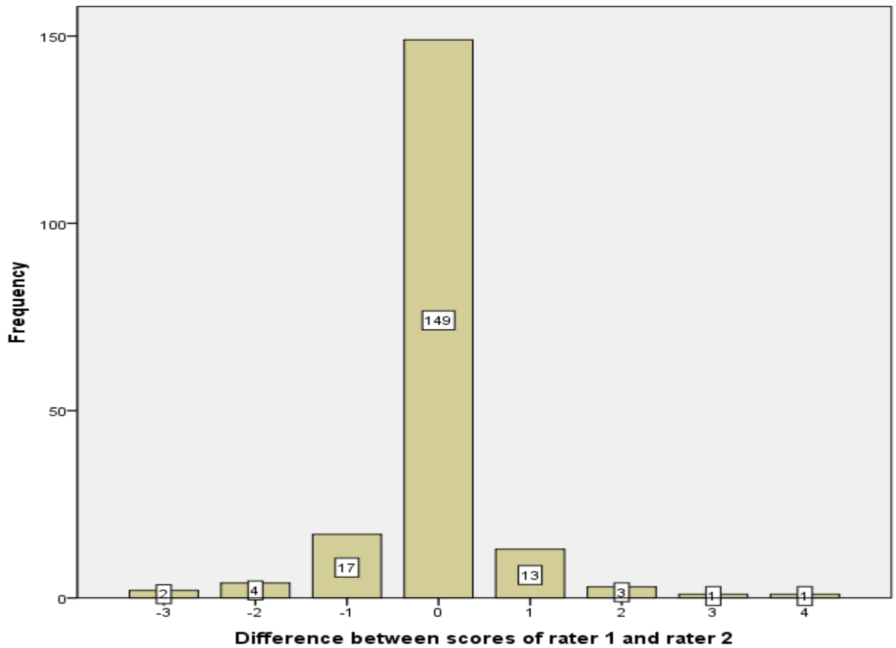


Figure 12 *Level of agreement between Rater 1 and Rater 2*

Although the result showed strong evidence that the scoring guides were being utilized in a consistent way, rater effect was examined to investigate whether scores were influened by a particular rater. Rater 1 (the one who received the training to score) was a more harsh rater than Rater 2 based on their estimates alone. The total error is 0.06+0.06=0.12. A 95% confidence interval for the mean mathematical discourse belief score between the two raters is -0.12 to 0.12. Since a difference between the two means includes 0, the two means are almost equal. Thus the difference between the beliefs scores is non-significant at α=0.05. Table 15 summarizes the result.

			Unweighted Fit			Weighted Fit		
	Estimate	Error	MNSQ	CI	T	MNSQ	CI	T
Rater 1	-0.02	0.06	0.79	(0.36, 1.64)	-0.6	0.84	(0.36, 1.64)	-0.4
Rater 2	0.02	0.06	1.13	(0.36, 1.64)	-1.3	0.62	(0.36, 1.64)	-1.3

Table 15 *Rater Effects*

It was clear that the scoring guide needs revision and that more training should be given to the raters. It appears that the description of each level needs to be clearer such that levels that are adjunct to each other are not overlapped in their descriptions.

Research Question 6 (RQ 6): To What Degree, if any, is Teacher Belief about Math Discourse Associated with Their Status?

There were two groups of respondents, preservice teachers and inservice teachers, in the study. Typically, t-test or ANOVA could be used to calculate the differences in teacher belief about mathematical discourse between the two groups, but these tests ignore measurement error. Thus latent regression was used to deal with the problem of measurement error. The motivation for the latent regression model arises from applications where distinct latent classes do not exist, but instead individuals vary according to a continuous latent variable (Tarpey & Petkova, 2008). Consider a simple linear multivariate regression model, where Y is the value of the dependent variable (belief about mathematical discourse), what is being predicted or explained, and β_0, a constant, equals the value of Y when the value of X=0, and β, the coefficient of X (status of respondent), ε is the error term.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon \qquad (5)$$

for $i = 1, 2, \ldots, n$.
When the predictor y is unobserved or latent in (1), we shall call the model a *latent regression model*, and our goal is to estimate the parameters of the model when y is unobserved.

Hypothesis

H0: There is no statistically significant difference in belief about mathematical discourse between preservice teachers and inservice teachers.

Ha: There is statistically significant difference in belief about mathematical discourse between preservice teachers and inservice teachers.

Preservice teachers were coded as 1, and inservice teachers were coded as 0. Based on the latent regression, the group difference is 0.10 logits, in which preservice teachers have a mean of 0.13 and inservice teachers have a mean of 0.07, where 0.10 is the grand mean of the person distribution. The total error is 0.03+0.03=0.06. A 95% confidence interval for the mean mathematical discourse belief score is -0.11 to 0.01. Since a difference between two means is 0, the hypothesis test proved that the means are equal. Table 16 summarizes the group difference.

			Unweighted Fit			Weighted Fit		
	Estimate	Error	MNSQ	CI	T	MNSQ	CI	T
Inservice teachers	-0.05	0.03	1.18	(0.49, 1.51)	0.7	1.17	(0.49, 1.51)	0.7
Preservice teachers	0.05	0.03	1.13	(0.76, 1.24)	1.1	1.14	(0.76, 1.24)	1.2

Table 16 *Latent Regression*

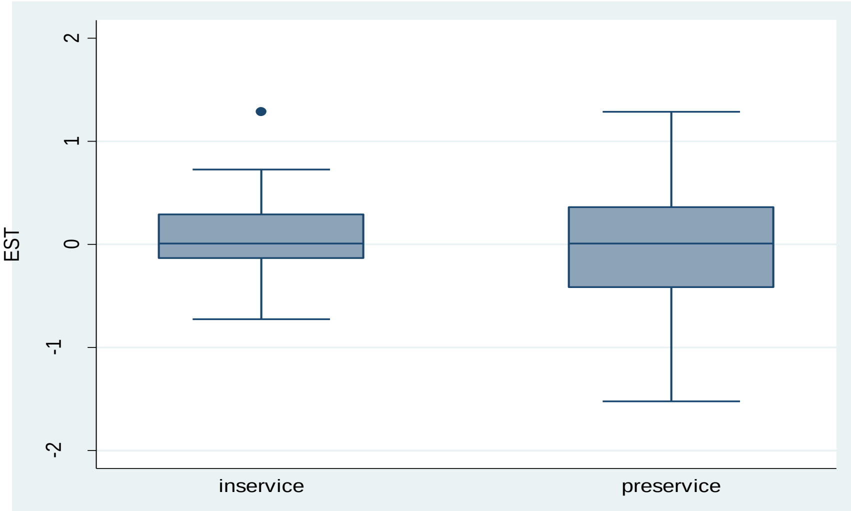


Figure 13 *Box plot of BMD scale scores (EAP) for inservice and preservice teachers*

As Figure 13 shows, the difference in belief about mathematical discourse between inservice and preservice teachers is negligible. In the sample of inservice teachers, there is an outlier. Looking at the location of the box within the whiskers, both samples are almost evenly distributed. The distribution of inservice teachers was slightly skewed right whereas the distribution of preservice teachers was rather symmetrical. The range of inservice teacher belief is narrower than that of preservice teachers. However, the two medians are almost equal.

CHAPTER 6: CONCLUSION

Statement of the Research Problem

This study addresses teachers' beliefs about pedagogy in mathematics reflected through discourse. Prior research indicates the importance of understanding teachers' beliefs, and investigates the links among teacher preparation context for learning, what preservice teachers learn, how their beliefs change, and how their learning are played out in practice in K-12 classrooms. Yet, we know little about these relationships. Assessing teacher beliefs has been difficult and often unsuccessful largely due to poor conceptualizations and measurement challenges associated with assessing beliefs. The field of teacher education is in need of carefully conceptualized and operationalized measures of teacher beliefs that are valid and reliable so that investigating relationships among teacher beliefs, classroom practice, and student outcomes is possible.

The purpose of this dissertation was to address the need for valid and reliable measures that can be used in the field of teacher education. In this dissertation, I developed an instrument to measure teachers' beliefs about mathematical discourse. The study was organized in nine phases using both qualitative and quantitative methods to gather validity and reliability evidence of the instrument. Although additional validity and reliability evidence is needed, the findings demonstrated that the BMD instrument is a potentially promising tool for informing and designing elementary mathematics method courses.

Review of the Theoretical Framework and the Methodology

I defined teacher beliefs about mathematical discourse as *the conscious and unconscious ideas and thoughts teachers have about how teachers and students should participate in classroom discussion to build mathematical knowledge.* My objective was to develop an instrument to measure teachers' beliefs along a continuum of beliefs about how teachers should interact with students in the process of developing mathematical ideas in the classroom. The theoretical framework of Scott and Mortimer (2005) was used to develop the construct theory. Central to this theoretical framework is the concept of a communicative approach that exists in two dimensions (the dialogic-authoritative dimension and the interactive-noninteractive dimension). Using this framework as a basis, the construct map was developed as a continuum to explore and describe qualitatively distinct levels of teacher beliefs about mathematical discourse namely, Univocal, Partial Univocal, Emerging Dialogical, and Dialogical (Figure 1).

Sample and Data Collection Procedure

Three purposive sample pools were developed to select participants from several populations: individuals who were enrolled in teacher educaton programs in the Spring semester of 2010, individuals who graduated from teacher education program in Fall 2009, and individuals who are currently teaching full-time in elementary classrooms. The survey was administered using Survey Monkey, and the data using the final instrument were collected from 202 respondents from the beginning of January to the beginning of Feburary, 2010. Out of 202 respondents, 82% (167) completed the survey.

Out of a total of 167 participants, 141 preservice teachers participated in the study from the California State University system, the University of California and one private university, and 27 inservice teachers from 16 states. The sample characteristics of this study reflect characteristics of California's teaching force in which teachers were predominantly white (70.1%) and female (72.4%). Preservice teachers came from diverse academic majors, and 48%

of them were in their first year in teacher education programs. Of the surveyed preservice teachers, 50% wished to teach in a suburban community after obtaining their teaching credential.

Instrumentation

The item design was based on the construct map (Figure 1), which was adapted from the theoretical framework of Scott and Mortimer (2005). The instrument that I designed featured 10 items using hypothetical teaching situations that present students' correct and incorrect thinking, and to then ask the teacher to respond in order to lead a mathematical discussion. Because beliefs tend to relate to specific contexts, the items were situated within contexts. Since beliefs dispose people toward particular actions, I provided respondents with opportunities to make teaching decisions. Based on a teacher's constructed responses, his or her belief was inferred. Mathematical content used in the items include patterns, addition, subtraction, multiplication with place values, base-10 block to represent whole numbers, the concept of multiple, sorting geometric figures based on their properties, and finding solutions for pairs of equations that are number sentences.

Summary of the Results

Research question 1: What is the evidence for content validity?

Evidence for construct validity consisted of a construct map, item design, outcome space, and measurement model (Wright map). According to the panel of experts including teacher educators, mathematics educators and master teachers, the BMD instrument's content coverage appeared relatively good.

Research question 2: What is the evidence for response process validity?

Analysis of post-hoc think-aloud interviews provided validity evidence for the response process. The respondents were generally positive about clarity of the BMD items; they reported that the item scenarios provoked their thinking about a teacher's action and how to best support students' mathematics learning. However, the interviews also revealed that the BMD measure lacked a clear definition of discussion (e.g., what counts as discussion and what does not), and I revised the items to strengthen the clarity of the definition of the construct.. The interviews also indicated that some BMD items lacked neutrality toward the topic of discussion, and these items were either eliminated or revised.

Research question 3: What is the evidence for internal structure validity?

Validity evidence for the internal structure of the measure was presented by showing that the empirical results of the Wright Map supported the theoretical framework in the construct map. The Wright Map (Wright & Masters, 1982) is a tool used to produce a visual representation of data (Wilson, 2004), and an empirically calibrated version of the construct map used to gather evidence based on internal structure of the construct that is hypothesized. Distinctions among the four levels of belief were reflected in the Wright Map as the theory suggested, although the distinction between level 3 (emerging dialogical) and level 4 (dialogical) was not as clear as the distinction between level 1 and level 2. Further research is needed to refine both the construct map and the items so the instrument has the capacity to measure a range of beliefs about how teachers should interact with students in the process of developing ideas in the classroom.

Research question 4: What is the evidence for external validity?

To address the validity evidence based on relations to variables external to the BMD measure, the study examined relationships between responses to the BMD measure and responses to the Preservice Teachers' Attitudes about Discourse in the Mathematics Classroom (PADM) measure (Casa, McGivney-Burelle, and DeFranco, 2008). Results indicated that the BMD measure and PADM measure delineated two orthogonal constructs. Post-hoc think-aloud interviews indicated that the PADM measure was useful for understanding teachers' general beliefs about mathematical discussion, while the BMD measure using specific teaching scenarios captures more complex beliefs about mathematical discussion. Instead of simply agreeing with the statement in the PADM measure that students should exchange their mathematical ideas with other students, the BMD measure provided respondents with the opportunity to articulate when mathematical discussion would be the most effective and appropriate during the instructional sequence. Moreover, another finding is that the PADM measure may be much more subject to social desirability bias than the BMD measure since 96% (157 out of 163) of participants showed higher scores of the PADM measure than the BMD measure, which suggests that teachers appear to be much more supportive of students engaging in mathematical discussion when their beliefs were assessed using the PADM measure than the BMD measure. Although, however, there are good theoretical reasons to believe that indirect questioning reduces social desirability bias, empirical research on the subject is limited and inconsluvie (Fisher, 1993).

Research question 5: What is the evidence for reliability?

As reliability evidence, Cronbach's alpha for the BMD was measured at 0.72, and Pearson Separation reliability was 0.75; both indicated a satisfactory level of reliability for the BMD measure. The standard error of measurement (SEM), the standard deviation of the estimated ability θ, was also examined to assess the accuracy of an estimate of a respondent's location since the SEM indicates how well a respondent is measured depending on her or his location to an item. SEM was consistent with the respondents' narrow distribution, indicating higher levels of error at the extreme end of the scale where there is generally less respondent information. Tthe BMD measure was not useful to measure respondents whose belief range falls between .5 logit to 1.0 logit. The inter-rater reliability for each item ranged from 0.50 to 0.97, a finding that suggests a need for more rater training.

Research question 6: To what degree, if any, are teacher beliefs about math discourse associated with other factors such as demographics, educational background and teaching experience?

The difference between inservice and preservice teachers in their beliefs about mathematical discourse was negligible. Looking at the location of the box within the whiskers, both samples were almost evenly distributed. The distribution of inservice teachers was slightly skewed right whereas the distribution of preservice teachers was rather symmetrical. The range of inservice teacher belief was narrower than that of preservice teachers. However, the two medians were almost equal.

Contributions of the Study

The purpose of this study was to develop and validate a measure of teaching beliefs based on psychometric analyses using Item Response Theory (IRT), which offers meaningful information about items and respondents, and supports interpretations of survey data. The findings from this study have implications for the field of mathematics education and teacher education. The BMD measure will be a useful tool to teacher educators, because it can be used to document what preservice teachers believe about mathematical discussion when they enter the program and to measure how their beliefs change as they progress through the program, complete the program, enter the teaching force, and continue in their teaching career. When teacher educators understand how teachers' beliefs change, as well as how their beliefs affect their teaching practice through the use of discussion (or vice versa), they can evaluate whether, and in what specific ways, their program can make an impact on preservice teachers' beliefs and teaching practices. Moreover, the BMD measure allows teacher educators to conceptualize teaching mathematics beyond using simple dichotomous camps but rather on a continuum.

Limitations and Further Research

This study has several limitations. The first pertains to assessing teachers' beliefs based on self-reported data. In this study, the respondents were asked to make teaching decisions from which I could infer beliefs. I recognized my reliance on inference throughout this process and categorized their responses. Although the items include contexts in which beliefs could emerge, I acknowledge that respondents' beliefs might look different given different scenarios. One of the strengths of the BMD instrument is to use learning episodes to create contexts to which respondents answer in their own words. This format generates qualitative data that can be used for multiple purposes. However, this strength came with a cost in terms of time required for raters to learn to use the scoring guide and translate the constructed responses into quantified responses. The alternative is to generate multiple-choice answers and see if the results prove to be similar of the constructed-response survey.

The second pertains to the limited audience for which the instrument was designed. The instrument was designed for prospective elementary school teachers, although it was piloted with practicing elementary school teachers as well. The pilot work using the survey with practicing elementary school teachers was promising in terms of its use as a measure of inservice teacherss' beliefs, but I make no claim about the efficacy of this survey with secondary school teachers.

The third limitation was sampling. The sampling of all applicable populations is important in order to insure that a measure is valid and reliable. Although this study included a substantial number of participants from teacher education programs, there are several populations that were not included, such as preservice teachers from private and independent universities or from any other alternative teacher preparation programs (e.g., Teach for America, Oakland Teaching Fellow). In addition, while this study was limited to the preservice teachers in California - mainly teacher education programs in California State University system and University of California system - inservice teachers were not limited to California; inservice teachers who participated in the survey were from all over the United States.

The fourth limitation of the study was a small number of participants (5 respondents) for think-aloud interviews. Selecting the sample for the think-aloud interviews was based on their voluntary participation, which may bring into question the representativeness of their responses. In addition, one respondent participated in think-aloud interviews three weeks after completing their survey whereas the rest of the interviewees participated in the think-aloud interviews within a week. This delay may have affected the responses due to memory loss. Future research should address the limitations, stated above, of this study. The sampling of all applicable populations is necessary in order to insure that a measure is valid and reliable. Preservice teachers from private and independent universities and alternative teacher preparation programs should be also recruited to participate in the study. In addition, more evidence for validity and reliability of the BMD measure should be gathered to strengthen the measure. This study was the first iteration of the instrument development process. The evidence for validity and reliability gathered in this study suggested that the instrument performed reasonably well. However, additional iterations are needed to gather more validity and reliability evidence so that the measure can be strengthened. Specifically, think-aloud interviews with a more substantial number of respondents, preferably immediately after completing the survey, can provide detailed information about the respondents' interpretations of the items they are asked.

Future research should also further investigate four levels of the construct map - namely Univocal, Partial Univocal, Emerging Dialogical, and Dialogical, or explore more levels in the continuum. Four levels of the construct map should be defined with clarity and described in a way that adjacent levels do not overlap. For example, one should ask; when does partial univocal discourse become emerging dialogical discourse? What types of teacher's questions reflect his or her beliefs that support partial univocal discourse, not emerging dialogical discourse?

Future research can also use the instrument, once strengthened and validated, in conjunction with an observational protocol. Convergent methods will enable investigations of relationships between teachers' manifested beliefs and their teaching practice.

Finally, future research can explore the relationship between teachers' knowledge of mathematics and teachers' beliefs about mathematical discourse. It is important to understand whether, or to what extent, preservice teachers' knowledge of mathematics affects their beliefs as well as the ways preservice teachers learn to create interactive classrooms in which mathematical communication is central.

Works Cited

Abelson, R. (1979) Differences between belief systems and knowledge systems. Cognitive Science, 3, 355-366.

Adams, R. J., & Khoo, S. T. (1996). *Quest*. Melbourne, Australia: ACER.

American Educational Research Association, American Psychological Association, National Council for Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Ambrose, R., Clement, L., Philipp, R., & Chauvot, J. (2004). Assessing prospective elementary school teachers' beliefs about mathematics and mathematics learning: Rationale and development of a constructed-response-format beliefs survey. *School Science and Mathematics, 104*(2), 56.

Anderson, J. C. (1978). The validity of Haire's shopping list projective technique, *Journal of Marketing Research*, 15, 644-649.

Andrich, D. A. (1978). A rating formulation for ordered response categories. *Psychometrika, 50,* 3-16.

Atkins, S. L. (1999). Listening to students: The power of mathematical conversations. *Teaching Children Mathematics,* 5(5), 289-295.

Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation. Retrieved January 27, 2006, from http://edresearch.org/irt/baker/final.pdf

Bakhtin, M. M. (1981). The dialogic imagination (Michael Holquist, Ed. And Caryl Emerson and Michael Holquist, Trans.). Austin: University of Texas Press.

Ball, D. L., Hill, H. C., & Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator,* 14-17, 20-22, 43-46.

Ball, D. L., & Rowan, B. (2004). Introduction: Measuring instruction. *The Elementary School Journal, 105* (1), 3-10.

Ball, D. (1993). With an eye toward the mathematical horizon: Dilemmas of teaching elementary school mathematics. *Elementary School Journal*, 93, 373–397.

Ball, D. L.. & Cohen, D. K. (1996). Reform by the book: what is – or might be – the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher, 25*, 6 - 8, 14.

Battista, M. T. (1993). Teacher beliefs and the reform movement in mathematics education. *Phi Delta Kappan, 75,* 462-470.

Bauersfeld, H. (1995). "Language games" in the mathematics classroom: Their function and their effects. In P. Cobb & H. Bauersfeld (Eds.), *The emergence of mathematical meaning: Interaction in classroom cultures* (pp. 271-289). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bearne, E. (1999). *Use of language across the secondary curriculum.* London: Routledge.

Belenky, M., Clinchy, B., Goldberger, N., & Tarule, J. (1986). *Women's ways of knowing: The development of self, voice and mind*. New York: Basic Books.

Benbow, R. M. (1993). Tracing mathematical beliefs of preservice teachers through integrated content-methods courses. *Proceedings of the Annual Conference of the American Educational Research Association.* (ERIC Document Reproduction Service No. ED 388638.)

Benbow, C. P., & Faulkner, L. R. (2008) Rejoinder to the critique of the National Mathematics Advisory Panel Final Report. *Educational Researcher*, 37(9), 645-648.

Ben-Peretz, M. (1990). The Teacher-Curriculum Encounter: Freeing Teachers from the Tyranny of Texts. New York: State University of New York Press.

Blanton, M.L. (2002). Using an undergraduate geometry course to challenge preservice teachers' notions of discourse. *Journal of Mathematics Teacher Education*, 5, 117–152.

Boaler, J. (1998). Open and closed mathematics: Student experiences and understandings. *Journal for Research in Mathematics Education*, 29(1), 41-62.

Boaler, J. (2008). When politics took the place of inquiry: A response to the National Mathematics Advisory Panel's Review of Instructional practices. *Educational Researcher*, *37*(9), 588–594.

Boaler, J. (2008). What's Math Got to Do With It? Helping Children Learn to Love Their Least Favorite Subject - and Why It's Important for America. (Viking/Penguin)

Boaler, J (2002). *Experiencing School Mathematics: Traditional and Reform Approaches to Teaching and their Impact on Student Learning.* Lawrence Erlbaum Associates: Mahwah, New Jersey.

Borko, H., Liston, D., & Whitcomb, J. A. (2007). Genres of empirical research in teacher education (Editorial). *Journal of Teacher Education, 58*(1), 3-11.

Brendefur, J., & Frykholm, J. (2000). Promoting mathematical communication in the classroom: Two preservice teachers' conception in the classroom: Two preservice teachers' conceptions and practices. *Journal of Mathematics Teacher Education*, 3(2), 125-153.

Britzman, DP. (1991). *Practice makes practice*. New York: State University of New York Press

Brooks, J. & Brooks, (1993). In Search of Understanding the Case for Constructivist Classrooms. Alexandria, VA: Association for Supervision and Curriculum Development.

Calder, B. J., & Robert E. B. (1977). Interpersonal influence on consumer behavior: an attribution theory approach, *Journal of Consumer Research*, 4, 29-38.

Carmines, E. G. & Zeller, R.A. (1991). *Reliability and validity assessment*. Newbury Park: Sage Publications.

Casa, T. M., McGivney-Burelle, J., & DeFranco, T. C. (2007). The Development of an Instrument to Measure Preservice Teachers' Attitudes about Discourse in the Mathematics Classroom, *School Science and Mathematics*, 107(2), 70-81.

Chazan, D. & Ball, D. L. (1999). Beyond being told not to tell. *For the Learning of Mathematics, 9*, 2-10.

Chapin, S. H, O'Connor, C & Anderson, N. C (2003). Classroom Discussions. Math Solutions. Sausalito, CA.

Cobb, P., Boufi, A., McClain, K. & Whitenack, J. (1997). Reflective discourse and collective reflection. Journal for Research in Mathematics Education, 17, 21–30.

Cobb, P., Wood, T, Yackel, E. & McNeal, B. (1992). Characteristics of classroom mathematics traditions: An interactional analysis. American Educational Research Journal, 29, 573–604.

Cochran-Smith, M., & Zeichner, K. (Eds.). (2005). *Studying teacher education: The report of the AERA panel on research and teacher education*. Mahwah, NJ: Lawrence Erlbaum Associates.

Confrey, J. (1990). What constructivism implies for teaching. In R. B. Davis, C. A. Maher, & N. Noddings (Eds.), *Constructivist views on the teaching and learning of mathematics* (pp. 107-124). Reston, CA: NCTM.

Cooney, T., Shealy, B., & Arvold, B. (1998). Conceptualizing belief structures preservice teachers secondary mathematics teachers. *Journal of Research in Mathematics Education*, 29(3), 306-333.

Cross, D. I. (2009). Alignment, cohesion, and change: Examining mathematics teachers' belief structures and their influence on instructional practices. *Journal of Mathematics Teacher Education*. 12, 325-346.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. Education Policy Analysis Archives, 8(1). Retrieved June 2002 from http://epaa.asu.edu/epaa/v8n1.

DeFranco, C. T., Gorgievski, N., & Truxaw, M. P. (2008). Measuring K-8 Teachers' Perceptions of Discourse Use in Their Mathematics Classes. *School Science Mathematics*. 108 (2), 55-70.

De Jong, Martijn G., Rik Pieters, and Jean-Paul Fox (2010), "Reducing Social Desirability Bias Through Item Randomized Response: An Application to Measure Underreported Desires," *Journal of Marketing Research*, 47 (February), 14–27.

Dewey, J. (1933). *How We Think, a Restatement of the Relation of Reflective Thinking to the Educative Process* (Revised ed.). Boston: D.C. Heath & Company.

Doerr, H. M., & Lesh, R. (2003). A modeling perspective on teacher development. In R. Lesh, & H. M. Doerr (Eds.), *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching.* (pp. 125-139). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Ebby, C.B. (2000). 'Learning to teach differently: The interaction between coursework and fieldwork for preservice teachers'. *Journal of Mathematics Teacher Education* 3, 69–97.

Edwards, A. L., & Thurstone, L. L. (1952). An interval consistency check for scale values determined by the method of successive intervals. Psychometrika,17,169-180.

Ernest, P. (1989). The knowledge, beliefs and attitudes of the mathematics teacher: A model. *Journal of Education for Teaching,* 15, 13-34.

Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 119-161). New York: Macmillian.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data.* Cambridge MA: MIT Press.

Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research, 38*(1), 47-64.

Fisher, R.J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research, Inc.* 20, 303-315.

Frykholm, J.A. (1996). Preservice teachers in mathematics: Struggling with the standards. Teaching and Teacher Education, 12(4), 665–681.

Gales, M. J., & Yan, W. (2001). Relationship between constructivist teacher beliefs and instructional practices to students' mathematical achievement: Evidence from TIMMS. Paper presented at the American Educational Research Association Annual Conference, Seattle, WA (ERIC Document Reproduction Service No. ED456133).

Gersten,R., Fuchs,L., Williams, J.P., & Baker, S. (2001). Teaching Reading Comprehension Strategies to Students With Learning Disabilities: Review of Research. *Review of Educational Research,* 71(2). 279–320.

Guha, R., Campbell, A., Humphrey, D., Shields, P., Tiffany-Morales, J., & Wechsler, M. (2006). *California's teaching force 2006: Key issues and trends*. Santa Cruz, CA: The Center for the Future of Teaching and Learning.

Goe, L., & Stickler, L.M. (2008, March). Teacher quality and student achievement: Making the most of recent research (TQ Research and Policy Brief). Washington, DC: National Comprehensive Center for Teacher Quality.

Good, T. L., & Grouws, D. A. (1977). Teaching effects: A process–product study in fourth-grade mathematics classrooms. *Journal of Teacher Education, 28*(3), 49–54.

Gonzales, P., Guzman, J. C., Partelow, L., Pahlke, E., Jocelyn, L., Kastberg, D., et al. (2004). *Highlights from the trends in international mathematics and science study (TIMSS), 2003. NCES 2005-005*ED Pubs, P.O. Box 1398, Jessup, MD 20794-7827. Tel: 877-433-7827 (Toll Free).

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Handal, B. (2003). Teachers' mathematical beliefs: a review. *The Mathematics Educator*, 13(2), 47-57.

Hiebert, J., Carpenter, T.P., Fennema, E., Fuson, K., Human, P. & Murray, H., et al.,(1996). Problem solving as a basis for reform in curriculum and instruction: The case of mathematics. Educational Researcher, 25(4), 12–21.

Hofer, B. (2001). Personal epistemology research: Implications for learning and teaching. Journal of Educational Psychology Review, 13, 353-383.

Hufferd-Ackles, K., Fuson, K. C., & Sherin, M. G. (2004). Describing levels and components of a math-talk learning community. *Journal for Research in Mathematics Education, 35*(2), 81–116.

Hungerford, T. W. (1994, January). Future elementary teachers: The neglected constituency. *American Mathematical Monthly*, 101, 15-21.

Kazemi, E. (1998). Discourse that promotes conceptual understanding. *Teaching Children Mathematics*, 4, 410-414.

Klatter E.B., Lodewijks H.G.L.C., Aarnoutse C.A.J. (2001). Learning conceptions of young students in the final year of primary education, *Learning and Instruction*, 11(6), 485-516.

Krummheuer, G. (1995). The ethnography of argumentation. In P. Cobb & H. Bauersfeld (Eds.), *The emergence of mathematical meaning: Interaction in classroom cultures* (pp. 229-269). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. American Educational Research Journal, 27, 29–63.

Lappan, G. (1997). The Challenges of implementation: Supporting teachers. *American Journal of Education*, 106(1). 207-239.

Laurenson, D, J. (1995). Mathematics and the drift towards conservatism: Are teacher beliefs and teaching practice following the beat of the same drummer? *National Consortium for Specialized Secondary Schools of Mathematics, Science, and Technology Journal*, 1(2), 3-7.

Lobato, J., Clarke, D., & Ellis, A. B. (2005). Initiating and eliciting in teaching: A reformulation of telling. *Journal for Research in Mathematics Education*, 36(2), 101-136.

Lobato, J. (2008) On learning processes and the National Mathematics Advisory Panel Report. *Educational Researcher*, 37(9), 595-601.

Lortie, D. (1975). *Schoolteacher: A sociological study.* Chicago: University of Chicago Press.

Manouchehri, A. (1997). School mathematics reform: Implications for mathematics teacher preparation. *Journal of Teacher Education*, 48, 197–209.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-74.

Mick, David G. (1996), "Are Studies of Dark Side Variables Confounded by Socially Desirable Responding? The Case of Materialism," *Journal of Consumer Research*, 23 (September), 106–119.

Monk, D.H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, 13(2), 125-145.

Morse, S. (1988, October). *Characteristics of migrant secondary students: Outreach meeting report*. (ERIC No. ED 319 550).

Mortimer, E. F., & Scott, P. H. (2003). Meaning making in secondary science classrooms. Maidenhead, UK: Open University Press.

Muis, K.R., Bendixen, L.D., & Haerle, F.C. (2006). Domain-generality and domain-specificity in personal epistemology research: Philosophical and empirical reflections in the development of a theoretical framework. Educational Psychology Review, 18, 3-54.

Murphy, P.K., Delli, L. A. M., & Edwards, M.N. (2004). The good teacher and good teaching: comparing beliefs of second-grade students, preservice teachers, and inservice teachers, *The Journal of Experimental Education* **72** (2), 69–92.

Nathan, M. J., & Knuth, E. J. (2003). A study of whole classroom mathematical discourse and teacher change. *Cognition and Instruction*, 21(2), 175-207.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Science Board, National Science Foundation, Division of Science Resources Statistics *Science and Engineering Indicators 2004*, Arlington, VA (NSB 04-01) [May 2004]

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

National Research Council. (2002). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, and B. Findell (Eds.). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Nespor, J. (1987). The role of beliefs in the practice of teaching. Journal of Curriculum Studies, 19, 317–328.

Ng, W., Nicholas, H., & Williams, A. (2010). School experience influences on pre-service teachers' evolving beliefs about effective teaching. Teaching and Teacher Education. 26(2). 278-289

Nisbert, S., & Warren, E. (2000). Primary school teachers' beliefs relating to mathematics teaching and assessing mathematics and factors that influence these beliefs. *Mathematics Education Research Journal*, 13(2), 34-47.

Office of Postsecondary Education (ED), Washington, DC. (2004). *Meeting the highly qualified teachers challenge: The secretary's third annual report on teacher quality*ED Pubs, P.O. Box 1398, Jessup, MD 20794-1398. Tel: 877-433-7827 (Toll Free).

Organisation for Economic Co-operation and Development (OECD)/Program for International Student Assessment (PISA). (2003). PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills.Paris, France: OECD.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62, 307-332.

Peressini, D. & Knuth, E.J. (1998). Why are you talking when you could be listening? The role of discourse and reflection in the professional development of a secondary mathematics teacher. Teaching and Teacher Education, 14(1), 107–125.

Philipp, R. (2007). Mathematics teachers' beliefs and affect. In F. K. Lester (Ed.), Second handbook of research on mathematics teaching and learning (pp. 257–315). Charlotte, NC: Information Age Publishing.

Philipp, R., Ambrose, R., Lamb, L., Sowder, J., Schappelle, B., Sowder, L., et al. (2007). Effects of early field experiences on the mathematical content knowledge and beliefs of prospective elementary school teachers: An experimental study. Journal for Research in Mathematics Education, 38(5), 438–476.

Raymond, A.M. (1997). Inconsistency between a beginning elementary school teacher's mathematics beliefs and teaching practice. Journal for Research in Mathematics Education, 28, 550–576.

Rimm-Kaufman, S., & Sawyer, B. (2004). Primary-grade teachers' self efficacy beliefs, attitudes toward teaching, and discipline and teaching practice priorities in relation to the responsive classroom approach. *The Elementary School Journal*, 104(2), 321-341.

Rokeach, M. (1968). Beliefs, attitudes, and values: *A theory of organization and change*. San Francisco: Jossey-Bass.

Rittenhouse, P. S. (1997). *Orienteering across the mathematical landscape: Using tasks and routines to foster student talk about mathematics*. Unpublished manuscript.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. Psychometrika Monograph Supplement.

Sarason, S. (1971). *The Culture of the School and the Problem of Change*. Boston: Allyn & Bacon.

Schoenfeld, A. (2004). The Math Wars. *Educational Policy,* 18(1), 253-286.

Schifter, D. (1996). *What's happening in math class? Reconstructing professional identities,* Vol. 2. New York: Teachers College Press.

Scott, P., & Mortimer, E. (2005). Meaning making in high school science classrooms: A framework for analyzing meaning making interactions. In K. Boersma, M. Goedhart, O. D. Jong, & H. Eijkelhof (Eds.), *Research and the quality of science education* (pp. 395–406). Netherlands: Springer.

Sherin, M. G. (2002). A balancing act: Developing a discourse community in a mathematics classroom. *Journal of Mathematics Teacher Education*, 5(3), 205-233.

Silver, E. A., & Smith, M. S. (1996). Building discourse communities in mathematics classrooms: A worthwhile but challenging journey. In P. C. Elliott (Ed.), *Communication in mathematics, K-12 and beyond: 1996 yearbook* (pp.20-28). Reston, VA: National Council of Teachers of Mathematics.

Steffe, L.P. & D'Ambrosio, B.S. (1995). Toward a working model of constructivist Teaching: A reaction to Simon. Journal for Research in Mathematics Education, 26, 114–145.

Stein, M. K., Engle, R. A., Hughes, E. K., & Smith, M. S. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. Source: *Mathematical thinking and learning* 10(4), 313-340.

Stiff, L. (2000). Constructivist mathematics and unicorns: National Council of Teachers of Mathematics president's message. Retrieved from http://www.nctm.org/news/president/2001-0708president.htm

Stigler, J., & Hiebert, J. (1997). Understanding and improving classroom mathematics, instruction: An overview of the TIMSS video study. *Phi Delta Kapgan*, Setptember 12-21.

Stipek, D. J., Givvin, K. B., Salmon, J. M., & MacGyvers, V. L., (2001). Teachers' beliefs and practices related to mathematics instruction, *Teaching and Teacher Education*, 17, 213-226.

Tarpey, T., & Petkova, E. (2009). Latent regression analysis. *Statistical Modelling*. To appear.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127-146). New York: Macmillan.

Torff, B., & Warburton, E. C. (2005). Assessment of teachers' beliefs about classroom use for critical thinking activities. *Educational and Psychological Measurement*, 65, 155-179.

Yackel, E. (1995). Children's talk in inquiry mathematics classrooms. In P. Cobb & H. Bauersfeld (Eds.), *The emergence of mathematical meaning: Interaction in classroom cultures* (pp. 131-162). Hillsdale, NJ: Lawrence Erlbaum Associates.

West, L. & Staub, F. C. (2003). *Content-Focused CoachingSM. Transforming mathematics lessons*. Portsmouth, NH: Heinemann.

Westfall, R. L., Harper W. B., & Donald T. C. (1957). The use of structured techniques in motivation research. *Journal of Marketing*, 22, 134-139.

Wertsch, J. V. & Toma, C. (1995). Discourse and learning in the classroom. In L.P. Steffe & J. Gale (Eds.), *Constructivism in education* (pp. 159-174). Hillsdale, NJ: Lawrence Erlbaum.

Wenglinsky, H. (2000). How teaching matters: Brining the classroom back into discussions of teacher quality. Princeton, NJ: Educational Testing Services.

Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach*. New Jersey: Lawrence Earlbaum Associates

Wilson, M., & Cooney, T. (2002). Mathematics teacher change and development. The role of beliefs. In G.C. Leder, E. Pehkonen, & G. To¨rner (Eds.), Beliefs: A hidden variable in mathematics education? (pp.127–147). Dordrecht: Kluwer Academic Publishers.

Wright, B. D., & Masters, G. N. (1981). *Rating scale analysis*. Chicago: MESA Press.

Wilcox, S., Schram, P., Lappan, G. & Lanier, P. (1991). The role of a learning community in changing preservice teachers' knowledge and beliefs about mathematics education. For the Learning of Mathematics, 11(3), 31–39.

Wu, M. L., Adams, R. J., & Wilson, M. (1998). ACER*ConQuest* [computer program]. Hawthorn, Australia: ACER.

Appendices

APPENDIX A: CONSTRUCT MAP OF TEACHERS' BELIEFS ABOUT MATH DISCOURSE

Respondents	Responses to the Items
4. Dialogical Teacher supports students to engage in conversation with one another about solutions, strategies, and to justify and argue with one another about their mathematical ideas in a whole-class format.	Responses indicate that teachers believe: "Students should share their answer with the class and discuss whether they agree or disagree with each other with explanations " "It's important to have students use their peers to work out the math problems in a whole-class discussion. They learn best from one another. I would have them challenge their own mathematical thinking and other's thinking by asking questions and making comments about multiple ways of solving the problem."
3. Emerging Dialogical Teacher elicits as many individual student's reasoning as possible. In the process, teacher encourages students to share their methods and thinking with one another. Teacher does not encourage students to ask questions, respond or make comments to other students' mathematical ideas.	Responses indicate that teachers believe: "It's important for the teacher to probe a student's thinking. I would ask students to come up with a different way to solve the problem." "I would survey how many different answers students came up with, and ask them to share with the class."
2. Partial Univocal Teacher focuses on conveying the exact meaning of math concepts by explaining and showing the correct methods to solve the problems. In the process, teacher leads students (either individually or collectively) through the IRE format to check their understanding and questions are corrective in nature. Teacher does neither encourage nor support students to share their mathematical thinking.	Responses indicate that teachers believe: "It is important for the teacher to go over step by step process to take the students to arrive at the correct answer." "I would walk a student through a problem to show the student the correct steps."
1. Univocal Teacher focuses on conveying the exact meaning of math concepts by explaining and showing the correct methods to solve the problems. Teacher focuses on correctness of student answers, and does not encourage students to articulate their methods.	Responses indicate that teachers believe: "I would point out that an answer is incorrect when a student gives an incorrect answer, and explain what was wrong with an answer." "Often times when students explain something, it can be a bit confusing so I try to repeat the explanation in my own words, pointing out the important parts of a student's answer."
N/A	Irrelevant responses "I would want my students to use manipulative".

APPENDIX B. BMD MEASURE (VERSION 1)

Directions: For each item, choose ONE answer that you feel is the MOST important.

[Context 1: After the discussion of the problem on the board, the teacher gives another multiplication problem (15 X 17) to the students to work in pairs. After students have worked on the problem in pairs, the teacher reconvenes the whole class for discussion.]

1. The teacher calls on a student to present his computation to the whole class and his solution is incorrect.

On board:

```
  15
X 17
----
 105
+ 15
----
 120
```

In this whole class discussion, the teacher should:

f. Model the correct method.
g. Ask questions to help him think through his reasoning.
h. Correct the student to help him understand the correct method.
i. Ask other students to comment and help the student think through his reasoning.
j. Other: ____

Are this question and response options clear to you? Yes ____ No _____
If not, please explain what is unclear to you:

2. The teacher calls on two more students (Jane, David) to present their computations on the board, and their solutions are correct.

Jane

```
   15
  X17
  ---
   35
   70
   50
+ 100
-----
  255
```

David

```
   15
 X 17
 ----
  105
+ 150
-----
  255
```

In this situation, the teacher should:

a. Move on to other problems.
b. Ask the students to explain how they got the answer.
c. Ask other students to make comments or ask questions.
d. Explain the students' methods so that everyone understands.
e. Other: ____

Are this question and response options clear to you? Yes ____ No _____
If not, please explain what is unclear to you:

[Context 2: The purpose of the problem is to introduce the idea of proof. After students have worked on a problem with their partner, the teacher reconvenes the whole class for discussion.]

Problem on the board:

Using nickels and dimes, find all the possible ways to make 20 cents (for example, 2 dimes=20 cents). Which way uses the most coins? Which way uses the least coins? How do you know that you have found all possible ways?

3. In this whole-class discussion, the teacher should:

a. Ask a student who knows the correct answer.
b. Ask students to share answers and ask others to comment.
c. Ask students to share any answer regardless of correctness.
d. Ask students who solved the task correctly to explain their method.
e. Other: ____

Are this question and response options clear to you? Yes ____ No _____
If not, please explain what is unclear to you:

[Context 3: The teacher asks students to work individually using base ten blocks to show the number 37 in more than one way.]

4. The teacher reconvenes the whole class for discussion. In the whole class discussion that followed, the teacher should:

a. Demonstrate grouping by tens using base 10 blocks.
b. Ask the students a series of questions to model grouping by ten.
c. Ask the students to share the block combinations they came up with.
d. Ask the students to talk to one another about how the different combinations are related.

e. Other: ____

Are this question and response options clear to you? Yes ____ No _____
If not, please explain what is unclear to you:

[Context 4: Students are sorting a set of polygons into two groups: triangles and other shapes. Within the set there are different sizes and types of triangles (scalene, equilateral, isosceles, and right triangles) and other polygons (squares, rectangles, hexagons, and parallelograms). After students have worked in small groups, the teacher reconvenes the class for discussion.]

5. The teacher noticed that one group of students has separated the equilateral triangles from all of the other triangles and grouped them alone. In this whole class discussion that followed, the teacher should:

   a. Point out the correct categorizations with the explanation.
   b. Review the properties of triangles and ask students to check their work.
   c. Tell that one group of students that they were incorrect in their categorization.
   d. Ask the students to explain their reasoning and have other students comment.
   e. Other: ____

Are this question and response options clear to you? Yes ____ No _____
If not, please explain what is unclear to you:

[Context 5: The purpose of the lesson is to introduce students to subtraction using decimals. After students have worked in small groups to discuss the task, the teacher reconvenes the class for discussion.]

<u>Problem on the board:</u>

Cathy had $25 and bought her favorite sweat shirt for $17.70. How much money did Cathy have left?

6. The teacher asks the students how to write a number sentence for their solution. The first student the teacher calls on gives a coherent explanation of the correct answer. In this whole-class discussion, the teacher should:

   a. Move on to other similar problems.
   b. Explain it so that everyone understands.

c. Ask the students to rephrase the explanation.
d. Repeat what the student said so that everyone can hear it.
e. Other: ____

Are this question and response options clear to you? Yes ____ No _____
If not, please explain what is unclear to you:

7. The teacher calls on two students (Juan and Sera) to present how they set up the subtraction problem.

Juan:	Sera:
25	25
-17.7	- 17.7

In this whole-class discussion, the teacher should:

a. Explain the mistake Juan made.
b. Ask Juan and Sera to explain their set up.
c. Tell the class which set up is correct to avoid confusion.
d. Ask the class to comment or ask questions of Juan or Sera.
e. Other: ____

Are this question and response options clear to you? Yes ____ No _____
If not, please explain what is unclear to you:

8. The teacher is introducing the mathematical term “even, and odd” when talking about numbers. For this whole-class discussion, the teacher should:

f. Provide students with a definition.
g. Chart students’ ideas as students share their definitions.
h. Ask students to share their definitions and invite others to comment.
i. Provide students with a definition of “even” and then ask them to guess what “odd” might be.
j. Other: ____

Are this question and response options clear to you? Yes ____ No _____
If not, please explain what is unclear to you:

[Context 6: Students are working with their seat partner to find solutions for a pair of equations; the equations are written as number sentences, and the unknowns are represented by squares and triangles. The teacher tells the class that the square and triangle in both equations stand for the same numbers.]

Problem on the board:

$$\square - \triangle = 15$$
$$\square + \triangle + \triangle = 21$$

9. During partner work, the teacher noticed that less than half the class found a solution that works for both equations. In this whole-class discussion that followed, the teacher should:

a. Ask those who got the answer to explain to the class.
b. Ask students to share answers, and then ask others to comment.
c. Explain the correct answer, since the majority did not get the answer.
d. Ask several students to share their answers (right or wrong) with the class.
e. Other: ____

Are this question and response options clear to you? Yes ____ No _____
If not, please explain what is unclear to you:

THE END (Thank you very much for your participation)

APPENDIX C: POST-HOC THINK-ALOUD PROTOCOL

To Participant:

In this interview I am interested in your thought process as you answer the survey. In order to do this I will ask you to talk aloud as you answer each question at a time. Please say out loud everything that comes into your mind while doing the task. If you come across any word that is unclear or confusing, please include that in this process as well.

Name: Date:

Starting time: Ending time:

APPENDIX D: BMD MEASURE (VERSION 2)

Dear prospective elementary teachers,

The purpose of this survey is to understand your opinions about mathematics teaching and your experience learning mathematics. Some of the questions to understand your opinions about mathematics teaching start with a teaching scenario and ask for your opinion about teaching strategies. There are no right or wrong answers to any of these questions. They ask you to explain your opinion; please provide clear and detailed answers.

The data gathered from this research will be used to develop an instrument. None of the data will be shared with your instructor, or used for the purpose of evaluating you, or passing judgment on your opinions. Your individual privacy will be maintained in all written or published data resulting from the study.

If you would like more information about this research, please feel free to contact me at (510) 512-8545 or heejujang@berkeley.edu. If you have questions about your rights as a study participant, you may contact the UC Berkeley's Office for the Protection of Human Subjects (OPHS) at (510) 885-4212 or ophs@berkeley.edu. The survey will take 15-20 minutes to complete. Thank you very much!

Sincerely,

Principal Investigator
Heeju Jang

Directions: For each question (1 to 7), please provide clear and detailed answers. The more descriptive and detailed your answers are, the better it is for me to understand your thinking. Remember that there is no right or wrong answers to any of these questions. The correctness of your answers is not the purpose of this survey.

1. A kindergarten teacher asks students to copy a pattern series and fill in the blank to continue the pattern.

<u>Problem on the board:</u>

△ □ ○ △ □ ○ △ □ ○ ___

The teacher reconvenes the class and asks Sarah to come up to the board to fill in the blank. Sarah fills the blank with a circle.

What should the teacher SAY and DO to guide students toward an understanding of the correct pattern?

2. A fourth grade teacher asks students to work individually to write a number sentence and solve the problem below. These students already learned about subtraction.

[Problem] You had $19.50 and bought a sweatshirt for $17.70. How much money did you have left?

When the teacher reconvenes the class, the first student the teacher calls on, Leandra, writes a correct number sentence and gives a coherent explanation.

In this situation, what should the teacher do next?

3. Second grade students are working on a subtraction problem. The teacher calls on Ben to present his computation to the class.

Below Ben shows his work.

$$\begin{array}{r} 12 \\ -\ 8 \\ \hline 16 \end{array}$$

In this situation, what should the teacher SAY and DO to guide students toward an understanding of subtraction?

4. A third grade teacher asks students to work individually on the problem 7 x 15. The teacher reconvenes the class and calls on three students to present their computations on the board.

Niral	Carlton	Jasmine
15	15	15
X 7	X 7	X 7
35	35	105
+ 7	+ 70	
42	105	

How should the teacher guide a discussion about the three computations?

5. A second grade teacher writes the problem below on the board, and asks students to work in pairs.

[Problem] Using nickels and pennies and dimes, find all the possible ways to make 10 cents.

1. Which way uses the most coins?
2. Which way uses the least coins?

When the teacher reconvenes the class, how should he/she lead a discussion of students' answers?

For each question (6 to 10), please choose ONE answer that best describes what you believe is MOST important. Remember that there are no right or wrong answers to any of these questions, and the correctness of your answers is not the purpose of this survey. For each question, please explain why your choice is most important as clearly as possible. The more detailed your answers are, the better it is for me to understand your thinking.

6. The mathematical term "multiples" was introduced to students in third grade. This year, their fourth grade teacher asks students to work individually to find multiples among numbers on the board:

8, 12, 16, 20, 24, 28, 32, 36

When the teacher reconvenes the class, what is MOST important for the teacher to do with the class?

a. Ask for answers and highlight correct answers while reviewing the definition of "multiple."

b. Ask students to agree or disagree with one another's ideas and explanations.

c. Guide students to identify multiples of 4, and then ask whether all numbers are also multiples of 8.

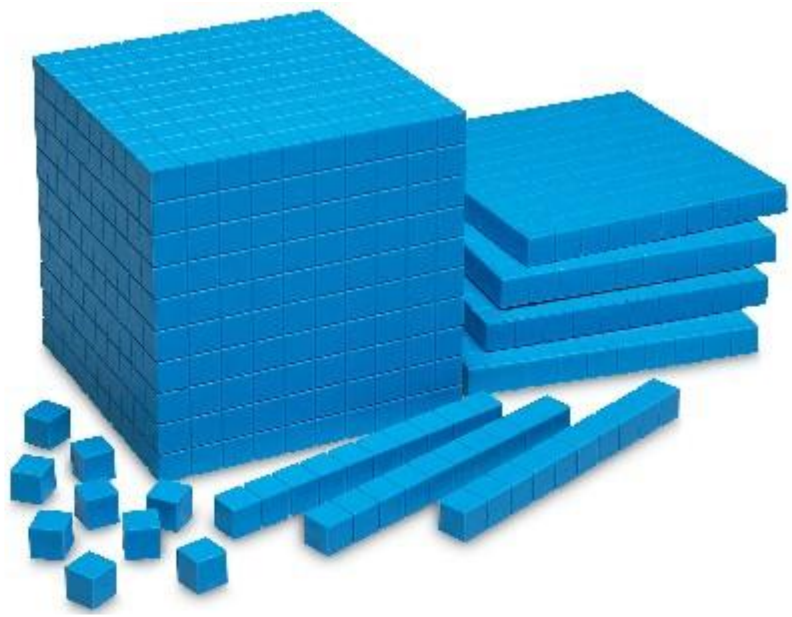
d. Elicit several students' ideas about multiples, asking questions to check for understanding.

e. Other (please specify)

Explain why your choice is most important.

7. Second grade students have been introduced to base ten blocks for the past few weeks. A second grade teacher asks students to work in pairs to show the number 406 with base ten blocks.

Base ten blocks



When the teacher reconvenes the class, what is the MOST important for the teacher to do with the class?

a. Ask several students to share and explain their block combinations.
b. Ask students leading questions to guide them to the correct model.
c. Guide students to comment on one another's block combinations and how they are related.
d. Demonstrate grouping by tens using base 10 blocks.
e. Other

Explain why your choice is most important.

8. A third grade teacher launches a geometry unit with the following sorting task.

[Problem] Sort your polygons into two groups: triangles and other shapes.

Students work in pairs. Their polygons include triangles (scalene, equilateral, isosceles, and right triangles) and other polygons (squares, rectangles, hexagons, and parallelograms). The teacher notices that some students sort into three groups: equilateral triangles, all other triangles, and all other polygons.

In this situation, the MOST important thing the teacher should do with the class is…

a. Ask students to share their groupings and ask questions to check for understanding.
b. Explain the properties of triangles and have students fix their errors.
c. Ask students to explain their sorts, and invite other students to ask the presenters questions and comment.
d. Ask questions about the properties of triangles to check for student understanding.
e. Other (please specify)

Explain why your choice is most important.

9. A fourth grade teacher calls on students to present how they set up the subtraction problem 25 – 17.7. Two students volunteer.

Juan	**Sera**
25	**25**
-17.7	**- 17.7**

In this situation, the MOST important thing the teacher should do with the class is…

a. Tell the class which set-up is correct.
b. Lead the class to comment or ask questions of Juan or Sera.
c. Explain the misconceptions about place value in Juan's set-up.
d. Ask if anyone set up the problem differently and to explain their reasoning.

10. A second grade teacher asks students to find solutions for a pair of equations. The teacher explains that the square and triangle in both equations stand for the same numbers.

$$\square - \triangle = 15$$
$$\square + \triangle + \triangle = 21$$

During partner work, the teacher notices that some students cannot find a solution that works for both equations.

In this situation, the MOST important thing the teacher should do with the class is…

a. Lead students to share their answers and make comments on one another's answers.

b. Re-explain the problem.
c. Ask several students to share their answers (right or wrong) and explain their reasoning.
d. Model another similar (easier) problem and then return to the problem.
e. Other (please specify)

Explain why your choice is most important.

Demographics

Gender: Male   Female

Age: Under 25   25–29   30–39   40–49   50–59   60 or older

Ethnicity:
African American
Asian
Caucasian
Hispanic
Pacific Islander
Native American
Other

Major(s) in undergraduate degree:

Please list MATH courses you have taken during your undergraduate study:

What kind of community do you currently teach?

__ Inner-city/Urban

__ Suburb

__ Rural

How many years have you taught?

__ Less than 1 year

__ 1-2 years

__ 3-4 years

__ 5-6 years

__ more than 7 years

What community would you like to teach after your graduation? Check all that apply.__ Inner-city/Urban

__ Suburb

__ Rural

APPENDIX E: CHANGES OF BMD ITEMS FROM VERSION 2 TO VERSION 3

Item	Math content in version 3	New item or modification?	Change in the item format (from multiple choice items to open-ended item)?
1	Pattern	New item	Yes
2, 3	Subtraction	Modification	
4	Multiplication and place values	Modification	
5	Addition	Modification: (Three pairs of students' work were added to the prompt).	
6	Using base-10 blocks to represent the whole number	Modification	
7	Subtraction and place values	Modification	
8	The concept of "multiple"	New item	No
9	Sorting geometric figures based on their properties	Modification	No
10	Finding solutions for pairs of equations that are number sentences	Modification	No

APPENDIX F: ITEM-SPECIFIC SCORING GUIDE FOR BMD VERSION 3

Item 1

Score	Description
4	The teacher asks Sarah to explain her thinking. The teacher encourages students to agree or disagree with Sarah's answer and explain why. The teacher encourages students to make comments on Sarah's answer or ask Sarah questions.
3	The teacher asks Sarah to explain her thinking. Followed by her response, the teacher asks if anyone else has the same or different answer, and asks them to share with the class.
2	The teacher points to each shape in the pattern with Sarah, and expects Sarah to catch her mistake in the process.
1	The teacher explains or shows why Sarah's answer is incorrect.
99	Irrelevant response. Nonsense.
.	Missing response

Item 2

Score	Description
4	The teacher asks the class if anyone has any questions about what Leandra did and if anyone agrees or disagrees with her answer and explanation. The teacher encourages students to restate or respond to Leandra's answer.
3	The teacher asks the class if anyone has different ideas on how to solve the problem, and invites them to share their answer and explanations with the class.
2	The teacher checks how many students got the correct answer, like Leandra.
1	The teacher praises Leandra on her correct answer (or acknowledges the correctness of her answer), and then the teacher moves on.
99	Irrelevant response. Nonsense.
.	Missing response

Item 3

Score	Description
4	The teacher asks students to agree or disagree with Ben's answer and to convince each other with explanations.
3	The teacher asks Ben to explain his thinking. The teacher also asks the class if anyone solved the problem differently and if they would like to share.
2	The teacher asks Ben to explain his thinking. The teacher leads Ben to the correct answer by asking step-by-step questions. The teacher asks Ben questions hoping that he will catch his own mistake.
1	The teacher says Ben's answer is incorrect and/or explains why

	and/or shows how to solve the problem. Or, the teacher is mainly concerned with the correctness of the answer.
99	Irrelevant response. Nonsense.
.	Missing response

Item 4

Score	Description
4	The teacher asks the class if they agree or disagree with any of the three computations and asks them to explain why. The teacher invites the students to make comments on the three computations or to help one another.
3	The teacher gives Niral, Carlton, and Jasmine the opportunity to explain their computations. The teacher encourages the students to share different methods of solving the problem.
2	The teacher gives Niral, Carlton, and Jasmine the opportunity to explain their computations. The teacher hopes Niral will catch his own mistake while he listens to Carlton and Jasmine explain their answers. The teacher asks Niral questions to arrive at the correct answer. The teacher focuses on Carlton's computation (the correct one).
1	The teacher tells the class which is the correct answer and/or explains the answer and/or shows how to solve the problem. Or the teacher reviews multiplication with the class.
99	Irrelevant response. Nonsense.
.	Missing response

Item 5

Score	Description
4	The teacher asks students to agree or disagree with the three answers and asks them to explain why.
3	The teacher asks the students to compare the number of coins used and share reasoning/explanations for their answer.
2	The teacher asks the three pairs to explain their answer. The teacher gives out hints or asks questions for the students to guide them in how to use fewer coins to make a dollar.
1	The teacher tells or shows the class how to use fewer coins to make a dollar.
99	Irrelevant response. Nonsense.
.	Missing response

Item 6

Score	Description
4	The teacher asks the students to share their way of representing the number 406 using base ten blocks with the class. And the teacher engages students to discuss the best way to use base ten blocks to represent 406.

3	The teacher asks the students to share multiple ways of representing the number 406 using base ten blocks with the class.
2	The teacher reviews what each cube, stick, and block represent, and/or asks students to explain their answer.
1	The teacher shows how to make 406 using base ten blocks.
99	Irrelevant response. Nonsense.
.	Missing response

Item 7

Score	Description
4	The teacher asks students to make comments on Juan's and Sera's answers. The teacher asks students to agree or disagree with Juan's and Sera's answers.
3	The teacher asks Juan question(s) to understand his thinking behind the incorrect subtraction set up, and asks if anyone else set up the subtraction problem differently from Juan or Sera. Or the teacher uses Juan's and Sera's answer as a way to start a discussion with the class. The teacher asks the differences between Juan's and Sera's set-up. The teacher asks the students about the differences between the two set-ups.
2	The teacher asks Juan and Sera to explain their answer. The teacher adds a decimal to show the alignment, and leads Juan to the correct set-up.
1	The teacher tells the class Sera's answer is correct (or Juan's answer is incorrect) and explains why, or reviews the place value, or shows how to correctly set up the subtraction problem (e.g., how to add a decimal and a zero to line up the numbers correctly).
99	Irrelevant response. Nonsense.
.	Missing response

Item 8-10

Score	Description
4	The teacher believes that it is important for students to engage in discussion where they agree or disagree with one another, and make comments or question questions to one another.
3	The teacher believes that it is important for students to listen to peers' different mathematical thinking and for them to share their thoughts with one another.
2	The teacher believes that it is important to ask questions of the students to guide them to the correct answer, or to help students to catch their own mistake.
1	The teacher believes that it is important to explain mathematical principles and show mathematical procedures.
99	Irrelevant response. Nonsense.
.	Missing response

APPENDIX G. BMD VERSION 3

Opinions about Mathematics Teaching

Dear prospective elementary teachers,

The purpose of this survey is to understand your opinions about mathematics teaching. The data from your responses will be used to develop a survey to be used by mathematics education instructors in future classes. None of the data will be shared with your instructor, or used for the purpose of evaluating you, or passing judgment on your opinions. Your individual privacy will be maintained in all written or published data resulting from the study.

If you would like more information about this research, please feel free to contact me at (510) 512-8545 or heejujang@berkeley.edu. If you have questions about your rights as a study participant, you may contact the UC Berkeley's Office for the Protection of Human Subjects (OPHS) at (510) 885-4212 or ophs@berkeley.edu. The survey will take 15-20 minutes to complete. Thank you very much!

Sincerely,

Heeju Jang
Principal Investigator

Directions: Questions 1-7 start with a teaching scenario and ask for your opinion about teaching strategies. There are no right or wrong answers to any of these questions. Please provide "clear and detailed" answers to help me to understand your opinions.

1. A kindergarten teacher asks students to copy a pattern series and fill in the blank to continue the pattern.

△□○△□○△□○___

The teacher reconvenes the class and asks Sarah to come up to the board to fill in the blank. Sarah fills the blank with a circle.

What should the teacher SAY and DO to guide students toward an understanding of the correct pattern?

Opinions about Mathematics Teaching

2. A fourth grade teacher asks students to work individually to write a number sentence and solve the problem below. These students already learned about subtraction.

[Problem] You had $19.50 and bought a sweatshirt for $17.70. How much money did you have left?

When the teacher reconvenes the class, the first student the teacher calls on, Leandra, writes a correct number sentence and gives a coherent explanation.

In this situation, what should the teacher do next?

3. Second grade students are working on a subtraction problem. The teacher calls on Ben to present his computation to the class.

Below Ben shows his work.

$$\begin{array}{r} 12 \\ -\ 8 \\ \hline 16 \end{array}$$

What should the teacher SAY and DO to guide students toward an understanding of subtraction?

Opinions about Mathematics Teaching

4. A third grade teacher asks students to work individually on the problem 7 x 15. The teacher reconvenes the class and calls on three students to present their computations on the board.

Niral	Carlton	Jasmine
15	15	15
X 7	X 7	X 7
35	35	105
+ 7	+70	
42	105	

How should the teacher guide a discussion about the three computations?

5. A second grade teacher gives each student pair 2 quarters, 5 dimes, 15 nickels and asks each pair to make one dollar using the least amount of coins.

Below students show their work.

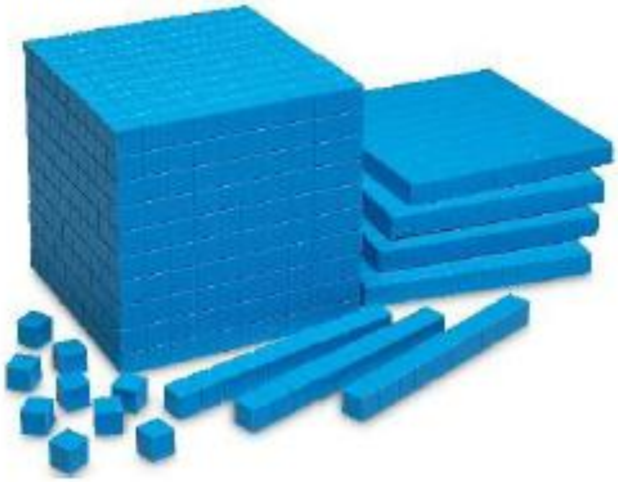
Pair 1: 2 quarters, 5 dimes
Pair 2: 1quarter, 15 nickels
Pair 3: 5 dimes, 10 nickels

When the teacher reconvenes the class, how should he/she lead a discussion of students' answers?

Opinions about Mathematics Teaching

6. A second grade students have been introduced to base ten blocks for the past few weeks. A second grade teacher asks students to work in pairs to show the number 406 with base ten blocks.

Base ten blocks



When the teacher reconvenes the class, how should he/she lead a discussion of students' answers?

7. A fourth grade teacher calls on students to present how they set up the subtraction problem 25 – 17.7. Two students volunteer.

Juan	Sara
25	25
-17.7	- 17.7

When the teacher reconvenes the class, how should he/she lead a discussion of students' answers?

Direction

Opinions about Mathematics Teaching

For Questions 8 - 10, please choose ONE answer that best describes what you believe is MOST important. For each question, please explain why your choice is most important as clearly as possible. Remember there are no right or wrong answers.

8. The mathematical term "multiples" was introduced to students in third grade. This year, their fourth grade teacher asks students to work individually to find multiples among numbers on the board:

8, 12, 16, 20, 24, 28, 32, 36

When the teacher reconvenes the class, what is MOST important for the teacher to do with the class?

- ◯ ask for answers, and highlight correct answers while reviewing the definition of 'multiple.'
- ◯ ask students to agree or disagree with one another's ideas and explanations.
- ◯ guide students to identify multiples of 4, and then ask whether all numbers are also multiples of 8.
- ◯ elicit several students' ideas about multiples, asking questions to check for understanding.
- ◯ Other

Other (please specify)

Explain why your choice is most important.

Opinions about Mathematics Teaching

9. A third grade teacher launches a geometry unit with the following sorting task.

[Problem] Sort your polygons into two groups: triangles and other shapes.

Students work in pairs. Their polygons include triangles (scalene, equilateral, isosceles, and right triangles) and other polygons (squares, rectangles, hexagons, and parallelograms). The teacher notices that some students sort into three groups: equilateral triangles, all other triangles, and all other polygons.

In this situation, the MOST important thing the teacher should do with the class is.....

- ◯ ask students to share their groupings and ask questions to check for understanding.
- ◯ explain the properties of triangles and have students fix their errors.
- ◯ ask students to explain their sorts, and invite other students to ask the presenters questions and comment
- ◯ ask questions about the properties of triangles to check for student understanding.
- ◯ Other

Other (please specify)

Explain why your choice is most important.

10. A second grade teacher asks students to find solutions for a pair of equations. The teacher explains that the square and triangle in both equations stand for the same numbers.

$$\square - \triangle = 15$$
$$\square + \triangle + \triangle = 21$$

Opinions about Mathematics Teaching

During partner work, the teacher notices that some students cannot find a solution that works for both equations.

In this situation, the MOST important thing the teacher should do with the class is.....

- ◯ lead students to share their answers and make comments on one another's answers.
- ◯ re-explain the problem.
- ◯ ask several students to share their answers (right or wrong) and explain their reasoning
- ◯ model another similar (easier) problem and then return to the problem.
- ◯ Other

Other (please specify)

Explain why your choice is most important.

About You

Information on your background will help me understand your opinions. All responses are confidential.

Gender

- ◯ Male
- ◯ Female

Age

- ◯ Under 25
- ◯ 25-29
- ◯ 30-39
- ◯ 40-49
- ◯ 50-59
- ◯ 60 or older

Opinions about Mathematics Teaching

Ethnicity

- [ ] African American
- [ ] Asian
- [ ] Caucasian
- [ ] Hispanic
- [ ] Pacific Islander
- [ ] Native American
- [ ] Other

Major(s) in undergraduate degree

Please list MATH courses you have taken during your undergraduate study.

What is your current status?

- ○ Recent graduate of teacher education program, but not have not entered teaching yet
- ○ in-service teacher
- ○ pre-service teacher

What kind of community do you currently teach?

- ○ Inner-city/Urban
- ○ Suburb
- ○ Rural

How many years have you taught?

- ○ Less than 1 year
- ○ 1-2 years
- ○ 3-4 years
- ○ 5-6 years
- ○ more than 7 years

Opinions about Mathematics Teaching

What grade level(s) do you currently teach?

In fulfilling your requirements to become a teacher in the State of California did you....

- ◯ Attend 1 year teacher preparation program
- ◯ Attend 2 year teacher preparation program
- ◯ Attend an alternative certification program (TFA, TOP, Teaching Fellows)

How would you characterize most of your class with respect to family income?

- ◯ Low family income
- ◯ Middle family income
- ◯ High family income

How many years of preservice teacher training did you have? Please round to the nearest whole number.

- ◯ 0 years
- ◯ 1 year
- ◯ 2 years
- ◯ 3 years
- ◯ 4 years
- ◯ 5 years

What kind of community would you like to teach? Choose all that apply.

- ☐ Inner-city/Urban
- ☐ Suburb
- ☐ Rural

What grade(s) would you like to teach?

Opinions about Mathematics Teaching

For the following statements, please indicate whether you agree or disagree.

1. Teachers should encourage debate about mathematical ideas.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

2. Mathematical tasks should encourage mathematical reasoning.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

3. Mathematical tasks should involve more than the memorization of facts.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

4. Mathematical tasks should be meaningful to students.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

Opinions about Mathematics Teaching

5. Teachers should encourage discussion in the mathematical classroom.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

6. Teachers and students should share their knowledge.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

7. Teachers should ask students questions that require higher-order thinking.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

8. All students should participate in mathematical discussions.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

Opinions about Mathematics Teaching

9. Teachers should probe students' understanding of mathematical ideas.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

10. Mathematical tasks should be complex, involving multiple sub-problems.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

11. Teachers should ask open-ended mathematical questions.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

12. Students should lead mathematical discussions.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

Opinions about Mathematics Teaching

13. Mathematical tasks should incorporate several mathematical questions.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

14. Mathematical tasks should be open-ended.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

15. Teachers should ask students for alternative strategies.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

16. Teachers should have students elaborate on their mathematical ideas.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

Opinions about Mathematics Teaching

17. Students should exchange their mathematical ideas with other students.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

18. Students' mathematical ideas should be valued.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

19. Teachers should allow ample time for students to grapple with mathematical ideas.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

20. Students should feel that their mathematical ideas are valued.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

Opinions about Mathematics Teaching

21. Students should listen to other students' mathematical ideas.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

22. Students should feel comfortable expressing their ideas in the classroom.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

23. Students should be a source of knowledge.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

24. Teachers should have students defend their mathematical ideas.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

Opinions about Mathematics Teaching

25. Mathematical tasks should encourage the use of different approaches.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

THANK YOU FOR YOUR PARTICIPATION!

APPENDIX H: ADDITIONAL DEMOGRAPHICS OF PARTICIPANTS

Major	**N**
Art	4
Business	7
Communication	8
Child Development	2
English	14
Elementary Education	11
Human Development	2
History	2
Science (e.g., Environmental Science, Biology)	4
Music	2
Sociology	12
Liberal Studies	65
Psychology	13
Math	2
Other*	20
Community they would like to teach	
Urban	29
Suburban	43
Rural	8
Urban or suburban	29
Suburban or rural	12
Urban, suburban, or rural	41
Community they are currently teaching**	
Urban	13
Suburban	17
Rural	7
Years of teaching experience**	
Less than 1 year	8
1-2 years	2
3-4 years	7
5-6 years	7
Over 7 years	13
Grade level currently teaching**	
1	3
2	9
3	4
4	1
5	5
Types of teacher preparation attended*	
1 year program	8

2 year program	9
Undergraduate teacher education program	13
State in which they are currently teaching**	**N**
AZ	1
CA	11
CO	1
CT	2
FL	2
GA	1
IL	1
IN	1
MA	1
MI	2
NE	1
NJ	1
NY	2
OH	1
PA	1
WA	2
WI	1

*Other includes agriculture science, American studies, bilingual education, comparative literature, criminal justice, economics, family studies, ethnic studies, German, international relations, library science, special education

**Inservice teachers only

APPENDIX I: WRIGHT MAP OF BMD MEASURE BY ITEM

Distribution of students	Distribution of Item									
	1	2	3	4	5	6	7	8	9	10
XX|										
|										
|										
2 |										
|										
|										
|										
|			3.4		5.4					
X|										
|										
|	1.4		3.3			6.4				
X|							7.4			
X|										10.4
|		2.4								
XX|	1.3						7.3			
1 XX|				4.4						
XX|								8.4		
XX|										
XXXX|										10.3
XXX|				4.3						
XXXXX|					5.3					
|									9.4	
XXXXX|								8.3		
XXXXXXXXXX|										
XXXX|										
|										
XXXXXX|										
0 XXXXXXXX|						6.3			9.3	
XXX|		2.3					7.2			
XXXX|										
XXXXX|				4.2						
X|										
XXXXXX|		2.2								
XXXXXXXXX|			3.2							
XX|										
XXX|										
|					5.2					
XXXXX|										
X|										
X|									9.2	10.2
-1 XXX|						6.2		8.2		
X|										
XXX|										
|										
X|										
|										
|									9.1	
X|	1.2							8.1		
X|										
|										
|										
X|										
-2 |										
|										
|					5.1					
|										
|							7.1			
X|	1.1	2.1	3.1	4.1		6.1				10.1

APPENDIX J: WRIGHT MAP OF BMD MEASURE – MULTIDIMENSIONAL ANALYSIS

```
       Dimension      Terms in the Model (excl Step terms)
--------------------
     Dimension Dimension 2              +item
-------------------------------------------------------------
                |         |                                  |
                |         |                                  |
   5            |        X|                                  |
                |         |                                  |
                |         |                                  |
                |         |                                  |
                |         |                                  |
                |         |                                  |
   4            |         |                                  |
                |        X|                                  |
                |         |                                  |
                |         |                                  |
                |         |                                  |
                |        X|                                  |
   3            |         |                                  |
                |         |                                  |
                |       XX|                                  |
                |         |                                  |
                |    XXXXX|                                  |
                |         |                                  |
   2            |       XX|                                  |
                |      XXX|                                  |
                |         |                                  |
                |        X|                                  |
              XX|     XXXX|                                  |
               X|       XX|                                  |
   1          XX|      XXX|                                  |
          XXXXXX|        X|                                  |
            XXXX|       XX|                                  |
        XXXXXXXX|      XXX|                                  |
             XXX|     XXXX|                                  |
      XXXXXXXXXX|      XXX|7                                 |
   0        XXXX|      XXX|3 5                               |
       XXXXXXXXX|  XXXXXXX|8                                 |
           XXXXX|       XX|4 6 9 10                          |
             XXX|      XXX|2 20 24                           |
            XXXX|     XXXX|1                                 |
  -1        XXXX|      XXX|23                                |
             XXX|     XXXX|34                                |
               X|       XX|22                                |
                |      XXX|11 21                             |
                |        X|16 18                             |
                |         |13 17 28                          |
  -2            |        X|12 14 15 19 25 26 27 29 31        |
                |         |30 33 35                          |
                |         |32                                |
                |         |                                  |
                |         |                                  |
                |         |                                  |
  -3            |         |                                  |
                |         |                                  |
                |         |                                  |
                |         |                                  |
                |         |                                  |
  -4            |         |                                  |
                |         |                                  |
                |         |                                  |
                |         |                                  |
                |         |                                  |
=============================================================
Each 'X' represents   2.4 cases
```

APPENDIX K: BMD ITEM ANALYSIS

```
item:1 (1)
Cases for this item    168   Discrimination  0.49
Item Threshold(s):    -3.82 -1.61  1.11  1.45   Weighted MNSQ   0.98
Item Delta(s):     -3.70 -1.70  2.28  0.24
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   WLEAvg:1 WLE SD:1
-----------------------------------------------------------------------------
   0       0.00        1       0.60   -0.11    -1.40(.164) -1.03     0.00
   1       1.00       25      14.88   -0.25    -3.36(.001) -0.41     0.60
   2       2.00      113      67.26   -0.16    -2.05(.042) -0.08     0.63
   3       3.00       13       7.74    0.23     3.03(.003)  0.53     0.38
   4       4.00       16       9.52    0.38     5.23(.000)  0.91     1.04
==============================================================================

item:2 (2)
Cases for this item    166   Discrimination  0.63
Item Threshold(s):    -3.14 -0.36 -0.02  1.08   Weighted MNSQ   0.99
Item Delta(s):     -3.12  0.71 -0.84  0.86
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   WLEAvg:1 WLE SD:1
------------------------------------------------------------------------------
   0       0.00        4       2.41   -0.18    -2.41(.017) -0.96     1.05
   1       1.00       56      33.73   -0.48    -6.99(.000) -0.47     0.51
   2       2.00       22      13.25   -0.10    -1.31(.193) -0.16     0.48
   3       3.00       53      31.93    0.26     3.39(.001)  0.26     0.49
   4       4.00       31      18.67    0.44     6.21(.000)  0.69     0.86
==============================================================================

item:3 (3)
Cases for this item    168   Discrimination  0.58
Item Threshold(s):    -2.95 -0.42  1.44  1.65   Weighted MNSQ   0.94
Item Delta(s):     -2.86 -0.45  2.90  0.06
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   WLEAvg:1 WLE SD:1
------------------------------------------------------------------------------
   0       0.00        5       2.98   -0.05    -0.66(.512) -0.17     0.31
   1       1.00       60      35.71   -0.42    -6.01(.000) -0.41     0.66
   2       2.00       87      51.79    0.12     1.60(.111)  0.10     0.51
   3       3.00        6       3.57    0.23     3.11(.002)  0.81     0.31
   4       4.00       10       5.95    0.45     6.44(.000)  1.39     1.04
==============================================================================

item:4 (4)
Cases for this item    167   Discrimination  0.65
Item Threshold(s):    -2.85 -0.18  0.66  0.95   Weighted MNSQ   1.00
Item Delta(s):     -2.80  0.09  1.47 -0.27
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   WLEAvg:1 WLE SD:1
------------------------------------------------------------------------------
   0       0.00        6       3.59   -0.15    -1.97(.051) -0.58     0.69
   1       1.00       66      39.52   -0.54    -8.29(.000) -0.48     0.53
   2       2.00       53      31.74    0.13     1.67(.097)  0.14     0.52
   3       3.00       14       8.38    0.18     2.34(.021)  0.41     0.33
   4       4.00       28      16.77    0.49     7.25(.000)  0.81     0.81
==============================================================================
```

item:5 (5)
Cases for this item 167 Discrimination 0.60
Item Threshold(s): -2.11 -0.74 0.61 1.67 Weighted MNSQ 0.94
Item Delta(s): -1.85 -0.80 0.79 1.30

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	9	5.39	-0.21	-2.72(.007)	-0.67	0.82
1	1.00	37	22.16	-0.46	-6.59(.000)	-0.61	0.55
2	2.00	70	41.92	0.05	0.64(.524)	0.07	0.52
3	3.00	36	21.56	0.26	3.47(.001)	0.36	0.50
4	4.00	15	8.98	0.37	5.04(.000)	0.91	0.99

item:6 (6)
Cases for this item 164 Discrimination 0.54
Item Threshold(s): -2.48 -0.97 0.12 1.41 Weighted MNSQ 1.00
Item Delta(s): -2.28 -0.89 0.09 1.15

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	5	3.05	-0.20	-2.58(.011)	-0.81	0.44
1	1.00	29	17.68	-0.36	-4.98(.000)	-0.57	0.59
2	2.00	55	33.54	-0.12	-1.53(.129)	-0.11	0.57
3	3.00	52	31.71	0.24	3.14(.002)	0.26	0.63
4	4.00	23	14.02	0.34	4.60(.000)	0.64	0.88

item:7 (7)
Cases for this item 168 Discrimination 0.58
Item Threshold(s): -2.30 -0.02 1.13 1.30 Weighted MNSQ 0.98
Item Delta(s): -2.20 0.05 2.60 -0.45

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	11	6.55	-0.08	-1.02(.311)	-0.25	0.90
1	1.00	71	42.26	-0.49	-7.25(.000)	-0.40	0.55
2	2.00	64	38.10	0.20	2.67(.008)	0.19	0.47
3	3.00	6	3.57	0.17	2.26(.025)	0.61	0.26
4	4.00	16	9.52	0.45	6.42(.000)	1.08	0.99

item:8 (8)
Cases for this item 164 Discrimination 0.52
Item Threshold(s): -1.56 -1.02 0.41 0.92 Weighted MNSQ 1.10
Item Delta(s): -0.75 -1.72 1.16 0.05

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	13	7.93	-0.25	-3.30(.001)	-0.59	0.39
1	1.00	17	10.37	-0.28	-3.77(.000)	-0.59	0.49
2	2.00	76	46.34	-0.09	-1.11(.267)	-0.06	0.62
3	3.00	25	15.24	0.06	0.81(.419)	0.12	0.44
4	4.00	33	20.12	0.44	6.18(.000)	0.65	0.88

```
item:9 (9)
Cases for this item    166   Discrimination  0.53
Item Threshold(s):    -1.54 -0.90  0.02  0.52   Weighted MNSQ   1.09
Item Delta(s):     -0.89 -1.29  0.55 -0.28
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   WLEAvg:1 WLE SD:1
------------------------------------------------------------------------------
   0       0.00       13       7.83   -0.21    -2.79(.006) -0.53     0.69
   1       1.00       19      11.45   -0.28    -3.72(.000) -0.55     0.61
   2       2.00       53      31.93   -0.21    -2.75(.007) -0.20     0.54
   3       3.00       30      18.07    0.08     0.97(.332)  0.12     0.59
   4       4.00       51      30.72    0.47     6.73(.000)  0.52     0.73
==============================================================================

Item 10
-------
item:10 (10)
Cases for this item    164   Discrimination  0.46
Item Threshold(s):    -2.41 -0.91  0.75  1.27   Weighted MNSQ   1.01
Item Delta(s):     -2.18 -1.05  1.47  0.41
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   WLEAvg:1 WLE SD:1
------------------------------------------------------------------------------
   0       0.00        6       3.66   -0.18    -2.33(.021) -0.68     0.75
   1       1.00       34      20.73   -0.18    -2.35(.020) -0.24     0.63
   2       2.00       82      50.00   -0.14    -1.82(.071) -0.08     0.59
   3       3.00       21      12.80    0.11     1.42(.157)  0.22     0.51
   4       4.00       21      12.80    0.42     5.92(.000)  0.84     0.97
==============================================================================

------------------------------------------------------------------------------
The following traditional statistics are only meaningful for complete
designs and when the amount of missing data is minimal.
In this analysis  1.07%  of the data are missing.

The following results are scaled to assume that a single response
was provided for each item.

N                               168
Mean                          21.09
Standard Deviation             5.82
Variance                      33.86
Skewness                       0.43
Kurtosis                       0.37
Standard error of mean         0.45
Standard error of measurement   3.09
Coefficient Alpha              0.72
==============================================================================
```

APPENDIX L: PADM ITEM ANALYSIS

item:1 (1)
Cases for this item 167 Discrimination 0.50
Item Threshold(s): -4.59 -3.29 -1.41 1.01 Weighted MNSQ 1.30
Item Delta(s): -4.30 -3.45 -1.45 0.92

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	1	0.60	0.03	0.36(.720)	0.54	0.00
1	1.00	4	2.40	-0.17	-2.16(.032)	-1.07	2.65
2	2.00	26	15.57	-0.33	-4.46(.000)	-0.82	1.07
3	3.00	76	45.51	-0.20	-2.62(.010)	-0.12	1.19
4	4.00	60	35.93	0.50	7.48(.000)	1.53	1.44

item:2 (2)
Cases for this item 165 Discrimination 0.72
Item Threshold(s): -6.25 -4.95 -3.08 -0.66 Weighted MNSQ 0.75
Item Delta(s): -5.96 -5.12 -3.12 -0.74

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	1	0.61	-0.45	-6.51(.000)	-5.15	0.00
2	2.00	1	0.61	-0.12	-1.56(.121)	-1.64	0.00
3	3.00	54	32.73	-0.53	-7.90(.000)	-0.79	0.75
4	4.00	109	66.06	0.62	9.97(.000)	1.03	1.39

item:3 (3)
Cases for this item 166 Discrimination 0.53
Item Threshold(s): -6.86 -5.56 -3.69 -1.27 Weighted MNSQ 1.95
Item Delta(s): -6.57 -5.72 -3.72 -1.35

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	3	1.81	-0.31	-4.14(.000)	-2.25	2.57
1	1.00	1	0.60	-0.11	-1.44(.153)	-1.64	0.00
2	2.00	2	1.20	-0.27	-3.63(.000)	-2.85	2.51
3	3.00	24	14.46	-0.31	-4.24(.000)	-0.87	0.89
4	4.00	136	81.93	0.49	7.26(.000)	0.67	1.46

item:4 (4)
Cases for this item 165 Discrimination 0.66
Item Threshold(s): -6.63 -5.33 -3.45 -1.03 Weighted MNSQ 0.93
Item Delta(s): -6.33 -5.49 -3.49 -1.12

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	1	0.61	-0.45	-6.45(.000)	-5.15	0.00
2	2.00	1	0.61	-0.19	-2.51(.013)	-2.48	0.00
3	3.00	42	25.45	-0.43	-6.08(.000)	-0.75	0.93
4	4.00	121	73.33	0.54	8.12(.000)	0.82	1.43

```
item:5 (5)
Cases for this item    166   Discrimination  0.67
Item Threshold(s):    -6.41 -5.12 -3.24 -0.82   Weighted MNSQ   0.97
Item Delta(s):     -6.12 -5.28 -3.28 -0.90
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   WLEAvg:1 WLE SD:1
------------------------------------------------------------------------------
   0       0.00        1       0.60   -0.43    -6.03(.000) -5.15     0.00
   2       2.00        4       2.41   -0.06    -0.76(.449) -0.44     0.73
   3       3.00       45      27.11   -0.55    -8.41(.000) -1.05     0.87
   4       4.00      116      69.88    0.62    10.21(.000)  0.95     1.39
==============================================================================

item:6 (6)
Cases for this item    162   Discrimination  0.74
Item Threshold(s):    -6.56 -5.27 -3.39 -0.98   Weighted MNSQ   0.80
Item Delta(s):     -6.28 -5.43 -3.43 -1.06
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   WLEAvg:1 WLE SD:1
------------------------------------------------------------------------------
   0       0.00        1       0.62   -0.46    -6.50(.000) -5.15     0.00
   2       2.00        1       0.62   -0.22    -2.84(.005) -2.73     0.00
   3       3.00       42      25.93   -0.52    -7.64(.000) -0.91     0.74
   4       4.00      118      72.84    0.63    10.23(.000)  0.94     1.37
==============================================================================


item:7 (7)
Cases for this item    166   Discrimination  0.63
Item Threshold(s):    -6.21 -4.91 -3.04 -0.62   Weighted MNSQ   1.12
Item Delta(s):     -5.92 -5.07 -3.07 -0.70
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   WLEAvg:1 WLE SD:1
------------------------------------------------------------------------------
   0       0.00        1       0.60   -0.43    -6.03(.000) -5.15     0.00
   2       2.00        7       4.22   -0.29    -3.93(.000) -1.57     0.67
   3       3.00       46      27.71   -0.33    -4.43(.000) -0.59     0.83
   4       4.00      112      67.47    0.51     7.57(.000)  0.90     1.52
==============================================================================


item:8 (8)
Cases for this item    166   Discrimination  0.65
Item Threshold(s):    -5.13 -3.83 -1.95  0.47   Weighted MNSQ   1.00
Item Delta(s):     -4.83 -3.99 -1.99  0.38
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   WLEAvg:1 WLE SD:1
------------------------------------------------------------------------------
   0       0.00        1       0.60   -0.45    -6.46(.000) -5.15     0.00
   1       1.00        1       0.60    0.00     0.01(.992)  0.07     0.00
   2       2.00       16       9.64   -0.31    -4.10(.000) -0.97     0.88
   3       3.00       72      43.37   -0.31    -4.12(.000) -0.29     0.98
   4       4.00       76      45.78    0.56     8.55(.000)  1.36     1.43
==============================================================================
```

item:9 (9)
Cases for this item 165 Discrimination 0.67
Item Threshold(s): -5.42 -4.13 -2.25 0.17 Weighted MNSQ 1.04
Item Delta(s): -5.13 -4.29 -2.29 0.09

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	1	0.61	-0.45	-6.47(.000)	-5.15	0.00
1	1.00	1	0.61	-0.09	-1.16(.246)	-1.27	0.00
2	2.00	11	6.67	-0.16	-2.08(.039)	-0.58	1.04
3	3.00	66	40.00	-0.46	-6.56(.000)	-0.55	0.90
4	4.00	86	52.12	0.61	9.90(.000)	1.30	1.37

item:10 (10)
Cases for this item 165 Discrimination 0.41
Item Threshold(s): -2.70 -1.41 0.47 2.89 Weighted MNSQ 1.44
Item Delta(s): -2.41 -1.57 0.43 2.80

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	1	0.61	-0.02	-0.22(.827)	-0.24	0.00
1	1.00	24	14.55	-0.17	-2.25(.025)	-0.33	1.13
2	2.00	71	43.03	-0.25	-3.31(.001)	-0.12	1.30
3	3.00	52	31.52	0.17	2.14(.034)	0.60	1.15
4	4.00	17	10.30	0.36	4.95(.000)	2.83	1.67

item:11 (11)
Cases for this item 165 Discrimination 0.72
Item Threshold(s): -4.31 -3.02 -1.14 1.28 Weighted MNSQ 0.88
Item Delta(s): -4.02 -3.18 -1.18 1.19

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	1	0.61	-0.45	-6.45(.000)	-5.15	0.00
1	1.00	3	1.82	-0.23	-3.02(.003)	-1.77	0.64
2	2.00	34	20.61	-0.40	-5.50(.000)	-0.80	0.89
3	3.00	75	45.45	-0.06	-0.71(.477)	0.07	0.89
4	4.00	52	31.52	0.55	8.32(.000)	1.82	1.42

item:12 (12)
Cases for this item 165 Discrimination 0.52
Item Threshold(s): -3.84 -2.55 -0.68 1.74 Weighted MNSQ 1.22
Item Delta(s): -3.56 -2.71 -0.71 1.66

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	1	0.61	-0.10	-1.26(.211)	-1.36	0.00
1	1.00	8	4.85	-0.24	-3.16(.002)	-1.11	0.96
2	2.00	36	21.82	-0.29	-3.90(.000)	-0.49	1.12
3	3.00	86	52.12	0.00	0.06(.954)	0.24	1.20
4	4.00	34	20.61	0.44	6.24(.000)	2.04	1.55

item:13 (13)
Cases for this item 165 Discrimination 0.45
Item Threshold(s): -3.34 -2.05 -0.17 2.24 Weighted MNSQ 1.20
Item Delta(s): -3.06 -2.21 -0.21 2.16

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	1	0.61	-0.02	-0.22(.829)	-0.24	0.00
1	1.00	8	4.85	-0.11	-1.40(.163)	-0.46	0.99
2	2.00	59	35.76	-0.32	-4.32(.000)	-0.35	1.10
3	3.00	75	45.45	0.06	0.74(.463)	0.35	1.25
4	4.00	22	13.33	0.44	6.26(.000)	2.76	1.38

item:14 (14)
Cases for this item 162 Discrimination 0.50
Item Threshold(s): -2.93 -1.63 0.25 2.66 Weighted MNSQ 0.96
Item Delta(s): -2.64 -1.79 0.21 2.58

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
1	1.00	14	8.64	-0.26	-3.45(.001)	-0.94	0.58
2	2.00	73	45.06	-0.26	-3.44(.001)	-0.20	1.05
3	3.00	61	37.65	0.20	2.53(.012)	0.67	1.37
4	4.00	14	8.64	0.39	5.37(.000)	3.26	1.20

item:15 (15)
Cases for this item 165 Discrimination 0.57
Item Threshold(s): -5.49 -4.20 -2.32 0.10 Weighted MNSQ 0.94
Item Delta(s): -5.20 -4.36 -2.36 0.02

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
1	1.00	1	0.61	-0.19	-2.52(.013)	-2.48	0.00
2	2.00	8	4.85	-0.14	-1.87(.064)	-0.60	1.20
3	3.00	73	44.24	-0.50	-7.29(.000)	-0.50	1.17
4	4.00	83	50.30	0.58	9.20(.000)	1.28	1.35

item:16 (16)
Cases for this item 163 Discrimination 0.76
Item Threshold(s): -5.85 -4.55 -2.68 -0.26 Weighted MNSQ 0.80
Item Delta(s): -5.56 -4.72 -2.72 -0.34

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	1	0.61	-0.45	-6.42(.000)	-5.15	0.00
1	1.00	1	0.61	-0.19	-2.51(.013)	-2.48	0.00
2	2.00	5	3.07	-0.25	-3.31(.001)	-1.32	1.73
3	3.00	58	35.58	-0.46	-6.52(.000)	-0.65	0.72
4	4.00	98	60.12	0.64	10.54(.000)	1.17	1.35

item:17 (17)
Cases for this item 164 Discrimination 0.75
Item Threshold(s): -5.59 -4.30 -2.42 0.00 Weighted MNSQ 0.75
Item Delta(s): -5.30 -4.46 -2.46 -0.09

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	1	0.61	-0.45	-6.44(.000)	-5.15	0.00
1	1.00	1	0.61	-0.19	-2.51(.013)	-2.48	0.00
2	2.00	7	4.27	-0.21	-2.73(.007)	-1.00	1.18
3	3.00	66	40.24	-0.47	-6.82(.000)	-0.60	0.78
4	4.00	89	54.27	0.65	10.92(.000)	1.29	1.34

item:18 (18)
Cases for this item 163 Discrimination 0.77
Item Threshold(s): -6.23 -4.93 -3.05 -0.63 Weighted MNSQ 0.79
Item Delta(s): -5.94 -5.09 -3.09 -0.72

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	1	0.61	-0.45	-6.47(.000)	-5.15	0.00
2	2.00	5	3.07	-0.30	-3.97(.000)	-1.75	0.75
3	3.00	47	28.83	-0.49	-7.20(.000)	-0.84	0.74
4	4.00	110	67.48	0.66	11.24(.000)	0.99	1.29

item:19 (19)
Cases for this item 166 Discrimination 0.71
Item Threshold(s): -5.72 -4.43 -2.55 -0.13 Weighted MNSQ 0.93
Item Delta(s): -5.43 -4.59 -2.59 -0.21

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	1	0.60	-0.43	-6.03(.000)	-5.15	0.00
1	1.00	3	1.81	-0.38	-5.30(.000)	-3.12	1.31
2	2.00	4	2.41	-0.14	-1.83(.069)	-0.98	1.29
3	3.00	63	37.95	-0.33	-4.50(.000)	-0.44	0.95
4	4.00	95	57.23	0.54	8.19(.000)	1.09	1.42

item:20 (20)
Cases for this item 165 Discrimination 0.78
Item Threshold(s): -6.23 -4.94 -3.06 -0.64 Weighted MNSQ 0.91
Item Delta(s): -5.94 -5.10 -3.10 -0.72

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	2	1.21	-0.53	-7.90(.000)	-4.89	0.37
1	1.00	1	0.61	-0.16	-2.08(.039)	-2.24	0.00
2	2.00	3	1.82	-0.23	-3.04(.003)	-1.88	0.79
3	3.00	46	27.88	-0.43	-6.11(.000)	-0.80	0.78
4	4.00	113	68.48	0.63	10.46(.000)	0.98	1.36

item:21 (21)
Cases for this item 165 Discrimination 0.63
Item Threshold(s): -5.80 -4.51 -2.63 -0.21 Weighted MNSQ 0.89
Item Delta(s): -5.51 -4.67 -2.67 -0.29

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
1	1.00	1	0.61	-0.17	-2.22(.028)	-2.24	0.00
2	2.00	8	4.85	-0.25	-3.31(.001)	-1.15	1.05
3	3.00	59	35.76	-0.51	-7.53(.000)	-0.69	0.97
4	4.00	97	58.79	0.63	10.39(.000)	1.18	1.36

item:22 (22)
Cases for this item 165 Discrimination 0.64
Item Threshold(s): -7.13 -5.84 -3.96 -1.54 Weighted MNSQ 1.28
Item Delta(s): -6.84 -6.00 -4.00 -1.62

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	1	0.61	-0.43	-6.04(.000)	-5.15	0.00
1	1.00	1	0.61	-0.32	-4.26(.000)	-4.62	0.00
2	2.00	3	1.82	-0.16	-2.07(.040)	-1.26	1.43
3	3.00	22	13.33	-0.32	-4.26(.000)	-0.93	0.70
4	4.00	138	83.64	0.50	7.46(.000)	0.67	1.46

item:23 (23)
Cases for this item 166 Discrimination 0.75
Item Threshold(s): -5.72 -4.43 -2.55 -0.13 Weighted MNSQ 0.94
Item Delta(s): -5.43 -4.59 -2.59 -0.21

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	1	0.60	-0.43	-6.03(.000)	-5.15	0.00
1	1.00	1	0.60	-0.32	-4.25(.000)	-4.62	0.00
2	2.00	12	7.23	-0.30	-3.98(.000)	-1.20	0.84
3	3.00	53	31.93	-0.37	-5.03(.000)	-0.59	0.87
4	4.00	99	59.64	0.62	10.15(.000)	1.14	1.36

item:24 (24)
Cases for this item 163 Discrimination 0.59
Item Threshold(s): -4.77 -3.47 -1.59 0.83 Weighted MNSQ 1.24
Item Delta(s): -4.48 -3.63 -1.63 0.74

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
1	1.00	4	2.45	-0.18	-2.34(.020)	-1.18	1.08
2	2.00	28	17.18	-0.35	-4.68(.000)	-0.77	0.90
3	3.00	60	36.81	-0.28	-3.75(.000)	-0.29	1.10
4	4.00	71	43.56	0.60	9.41(.000)	1.45	1.31

item:25 (25)
Cases for this item 166 Discrimination 0.77
Item Threshold(s): -6.13 -4.84 -2.96 -0.54 Weighted MNSQ 0.80
Item Delta(s): -5.84 -5.00 -3.00 -0.62

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1
0	0.00	1	0.60	-0.43	-6.03(.000)	-5.15	0.00
1	1.00	1	0.60	-0.32	-4.25(.000)	-4.62	0.00
2	2.00	5	3.01	-0.25	-3.35(.001)	-1.61	0.69
3	3.00	50	30.12	-0.44	-6.19(.000)	-0.77	0.67
4	4.00	109	65.66	0.63	10.46(.000)	1.05	1.38

The following traditional statistics are only meaningful for complete designs and when the amount of missing data is minimal.
In this analysis 1.29% of the data are missing.

The following results are scaled to assume that a single response was provided for each item.

N	167
Mean	83.99
Standard Deviation	11.16
Variance	124.45
Skewness	-1.56
Kurtosis	5.41
Standard error of mean	0.86
Standard error of measurement	3.72
Coefficient Alpha	0.89

APPENDIX M: PADM SUB-SCALES

Component 1: Promoting mathematical reasoning

3. Teachers should encourage debate about mathematical ideas.
6. Mathematical tasks should encourage mathematical reasoning.
26. Mathematical tasks should involve more than the memorization of facts.
8. Mathematical tasks should be meaningful to students.
7. Teachers should encourage discussion in the mathematics classroom.
4. Teachers and students should share their knowledge.
10. Teachers should ask students questions that require higher-order thinking.
1. All students should participate in mathematical discussions.
11. Teachers should probe students' understanding of mathematical ideas.

Component 2: Examining complex mathematical concepts

16. Mathematical tasks should be complex, involving multiple sub-problems.
25. Teachers should ask open-ended mathematical questions.
12. Students should lead mathematical discussions.
24. Mathematical tasks should incorporate several mathematical questions.
1. Mathematical tasks should be open-ended.
23. Teachers should ask students for alternative strategies.
21. Teachers should have students elaborate on their mathematical ideas.

Component 3: Valuing students' mathematical ideas

18. Students should exchange their mathematical ideas with other students.
20. Students' mathematical ideas should be valued
17. Teachers should allow ample time for students to grapple with mathematical ideas.
9. Students should feel that their mathematical ideas are valued.
5. Students should listen to other students' mathematical ideas.
22. Students should feel comfortable expressing their ideas in the classroom.
15. Students should be a source of knowledge.
13. Teachers should have students defend their mathematical ideas.
14. Mathematical tasks should encourage the use of different approaches.

APPENDIX N: SELECTED PADM ITEMS

Teachers should encourage debate about mathematical ideas.
Teachers should encourage discussion in the mathematical classroom.
Teachers and students should share their knowledge.
All students should participate in mathematical discussions.
Teachers should probe students' understanding of mathematical ideas.
Students should lead mathematical discussions.
Teachers should have students elaborate on their mathematical ideas.
Students should exchange their mathematical ideas with other students.
Students should listen to other students' mathematical ideas.
Students should feel comfortable expressing their ideas in the classroom.
Students should be a source of knowledge.
Teachers should have students defend their mathematical ideas.

APPENDIX O: DESCRIPTIVE STATISTICS OF PADM ITEMS

item	N	Minimum	Maximum	Mean	Std. Deviation	% respondents SA/A	Neutral	D/DS
S1	166	0	4	3.15	0.79	136	26	4
S2	165	0	4	3.64	0.56	163	1	1
S3	165	0	4	3.75	0.68	160	1	4
S4	165	0	4	3.71	0.54	163	1	1
S5	165	0	4	3.66	0.59	160	4	1
S6	162	0	4	3.70	0.54	160	1	1
S7	165	0	4	3.61	0.63	157	7	1
S8	166	0	4	3.33	0.73	148	16	2
S9	165	0	4	3.42	0.70	152	11	2
S10	165	0	4	2.36	0.88	69	71	25
S11	165	0	4	3.05	0.81	127	34	4
S12	165	0	4	2.87	0.81	120	36	9
S13	165	0	4	2.66	0.79	197	59	9
S14	162	1	4	2.46	0.77	75	73	14
S15	165	1	4	3.44	0.62	156	8	1
S16	163	0	4	3.54	0.65	156	5	2
S17	164	0	4	3.47	0.67	157	7	2
S18	163	0	4	3.63	0.61	157	5	1
S19	165	0	4	3.51	0.67	158	4	3
S20	164	0	4	3.64	0.62	159	3	2
S21	165	1	4	3.53	0.62	156	8	1
S22	164	0	4	3.80	0.52	160	3	1
S23	165	0	4	3.51	0.69	152	12	1
S24	163	1	4	3.21	0.81	131	28	4
S25	165	0	4	3.61	0.61	159	5	1

APPENDIX P. AVERAGE SCORES OF BMD MEASURE AND PADM MEASURE

ID	Average Score of BMD Measure	Average Score of PADM Measure	Difference
98	3.9	3.5	-0.4
195	4.0	3.7	-0.3
136	2.5	2.2	-0.3
36	1.2	0.9	-0.3
95	3.0	2.8	-0.2
39	3.3	3.2	-0.1
91	2.6	2.6	0.0
205	2.2	2.3	0.1
57	3.1	3.2	0.1
150	2.6	2.8	0.2
187	3.0	3.2	0.2
186	3.8	4.0	0.2
198	2.9	3.1	0.2
49	3.0	3.2	0.2
142	2.6	2.9	0.3
85	2.6	2.9	0.3
58	2.5	2.8	0.3
69	2.1	2.4	0.3
194	2.5	2.8	0.3
179	2.9	3.4	0.5
155	2.6	3.1	0.5
42	3.1	3.6	0.5
224	3.3	3.8	0.5
161	2.8	3.3	0.5
35	2.3	2.8	0.5
104	2.4	3.0	0.6
143	2.1	2.7	0.6
108	2.4	3.0	0.6
62	2.3	2.9	0.6
96	2.1	2.7	0.6
141	3.2	3.8	0.6
188	2.8	3.5	0.7
223	2.5	3.2	0.7
37	3.2	3.9	0.7
159	2.0	2.7	0.7
55	2.7	3.4	0.7
165	2.7	3.4	0.7

178	2.1	2.8	0.7
52	2.2	3.0	0.8
125	2.3	3.1	0.8
227	2.4	3.2	0.8
50	2.7	3.5	0.8
81	2.4	3.2	0.8
157	2.4	3.2	0.8
116	2.6	3.5	0.9
180	2.8	3.7	0.9
43	2.7	3.6	0.9
183	1.5	2.4	0.9
239	2.1	3.0	0.9
218	2.6	3.5	0.9
229	2.4	3.3	0.9
176	2.8	3.7	0.9
76	2.4	3.4	1.0
92	2.1	3.1	1.0
169	2.5	3.5	1.0
170	1.9	2.9	1.0
67	1.7	2.7	1.0
79	2.0	3.0	1.0
156	2.4	3.4	1.0
204	2.4	3.4	1.0
109	2.7	3.7	1.0
233	1.7	2.7	1.0
160	1.8	2.8	1.0
63	2.2	3.2	1.0
78	2.8	3.8	1.0
86	2.4	3.4	1.0
113	2.2	3.2	1.0
149	2.6	3.6	1.0
166	2.8	3.8	1.0
68	2.6	3.7	1.1
118	2.8	3.9	1.1
206	2.1	3.2	1.1
84	2.2	3.3	1.1
99	2.7	3.8	1.1
181	2.1	3.2	1.1
222	1.8	3.0	1.2
139	2.0	3.2	1.2
44	2.1	3.3	1.2

38	2.6	3.8	1.2
184	1.8	3.0	1.2
202	1.8	3.0	1.2
117	2.2	3.4	1.2
145	2.4	3.6	1.2
152	2.0	3.2	1.2
131	2.5	3.8	1.3
56	2.2	3.5	1.3
124	1.9	3.2	1.3
51	2.2	3.5	1.3
41	1.8	3.1	1.3
127	2.3	3.6	1.3
138	2.3	3.6	1.3
122	1.9	3.3	1.4
201	2.2	3.6	1.4
59	1.9	3.3	1.4
123	2.1	3.5	1.4
158	1.3	2.7	1.4
154	2.4	3.8	1.4
213	1.8	3.3	1.5
148	2.1	3.6	1.5
189	2.3	3.8	1.5
100	1.8	3.3	1.5
217	1.8	3.3	1.5
210	1.3	2.8	1.5
71	1.5	3.0	1.5
173	1.7	3.2	1.5
242	2.2	3.8	1.6
107	1.9	3.5	1.6
65	1.8	3.4	1.6
215	1.4	3.0	1.6
167	1.6	3.2	1.6
219	1.4	3.0	1.6
175	2.1	3.8	1.7
46	1.7	3.4	1.7
120	1.7	3.4	1.7
47	1.9	3.6	1.7
228	2.3	4.0	1.7
64	1.8	3.5	1.7
235	1.2	2.9	1.7
72	1.9	3.6	1.7

240	1.6	3.3	1.7
153	2.1	3.8	1.7
146	2.0	3.8	1.8
162	2.0	3.8	1.8
171	1.6	3.4	1.8
225	2.2	4.0	1.8
208	1.5	3.3	1.8
74	1.9	3.7	1.8
103	1.2	3.0	1.8
234	1.5	3.3	1.8
48	2.0	3.8	1.8
119	2.0	3.8	1.8
191	1.6	3.4	1.8
82	1.7	3.6	1.9
61	1.8	3.7	1.9
203	1.8	3.7	1.9
94	1.9	3.8	1.9
192	1.7	3.6	1.9
89	1.5	3.4	1.9
193	1.5	3.4	1.9
231	1.9	3.8	1.9
209	1.5	3.5	2.0
132	1.7	3.7	2.0
207	1.3	3.3	2.0
182	1.6	3.6	2.0
216	1.1	3.2	2.1
102	1.8	3.9	2.1
106	1.7	3.8	2.1
126	1.7	3.8	2.1
54	1.5	3.7	2.2
230	1.2	3.4	2.2
199	1.5	3.8	2.3
212	1.7	4.0	2.3
177	1.5	3.8	2.3
70	1.5	3.9	2.4
128	1.3	3.7	2.4
226	0.9	3.3	2.4
220	1.2	3.6	2.4
190	0.8	3.2	2.4
112	1.3	3.8	2.5
130	1.3	3.9	2.6

214	1.2	3.8	2.6
232	1.1	3.8	2.7
66	0.7	3.6	2.9

APPENDIX Q: WRIGHT MAP FOR PADM MEASURE

```
                                +item
----------------------------------------------------
   6            |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
   5            |                                  |
             XXX|                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
   4            |                                  |
              XX|                                  |
                |                                  |
               X|                                  |
                |                                  |
   3           X|                                  |
                |                                  |
              XX|                                  |
                |                                  |
         XXXXXXX|                                  |
   2        XXXX|                                  |
                |                                  |
           XXXXX|                                  |
              XX|                                  |
         XXXXXXX|                                  |
   1         XXX|                                  |
           XXXXX|                                  |
           XXXXX|                                  |
          XXXXXX|                                  |
          XXXXXX|                                  |
   0      XXXXXX|                                  |
      XXXXXXXXXX|10                                |
         XXXXXXX|14                                |
           XXXXX|                                  |
            XXXX|13                                |
  -1    XXXXXXXX|                                  |
            XXXX|                                  |
          XXXXXX|12                                |
               X|                                  |
            XXXX|11                                |
               X|                                  |
  -2            |1                                 |
               X|24                                |
               X|                                  |
               X|8                                 |
                |9                                 |
  -3            |15 17                             |
                |16 19 21 23                       |
                |                                  |
                |2 7 18 20 25                      |
                |5                                 |
  -4            |4 6                               |
                |                                  |
                |3                                 |
               X|22                                |
                |                                  |
  -5            |                                  |
               X|                                  |
                |                                  |
====================================================
Each 'X' represents   1.4 cases
```