DEVELOPING AND VALIDATING A NEW INSTRUMENT TO MEASURE THE SELF-EFFICACY OF ELEMENTARY MATHEMATICS TEACHERS

by

Jennifer Richardson McGee

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Education in
Educational Leadership

Charlotte

2012

Approved by:

____
Dr. Chuang Wang

____
Dr. Rich Lambert

____
Dr. James J. Bird

____
Dr. Andrew B. Polly

ABSTRACT

JENNIFER RICHARDSON MCGEE. Developing and validating a new instrument to measure the self-efficacy of elementary mathematics teachers. (Under direction of Dr. CHUANG WANG)

The purpose of this study was the development and validation of an instrument to measure the self-efficacy of elementary mathematics teachers. Self-efficacy, as defined by Bandura, was the theoretical framework for the development of the instrument. The complex belief systems of mathematics teachers, as touted by Ernest (1989) provided insight into the elements of mathematics beliefs that could be relative to a teacher's self-efficacy beliefs. The Self-efficacy for Teaching Mathematics Instrument (SETMI) was developed in August 2010 and has undergone revisions to the original version through processes defined in this study. Evidence of reliability and validity were collected to determine if the SETMI is an adequate instrument to measure self-efficacy of elementary mathematics teachers. Findings indicate that reliability of the instrument is adequate but that the original constructs represented by the instrument may be different than initially anticipated. Construct validity of the revised SETMI was tested using Confirmatory Factor Analysis and while the measurement models fit the covariance matrix, the items that represent mathematics content may in fact be better measures of self-concept for mathematics content knowledge. These findings indicate evidence of a potential structural model for self-efficacy.

ACKNOWLEDGMENT

The completion of this dissertation and degree would not have been possible without the contribution of several important individuals. First, my wonderfully patient husband Jimmy has been a source of strength and encouragement to begin and complete this adventure. My son Carter, who is only11 months old at the completion of this program, has been a wonderful addition to our lives and is a constant source of motivation. Thank you both for being patient with me when I had to have alone time to write, think, and reflect. I am also extremely fortunate to have a dissertation committee that consists of colleagues, mentors, and friends. First and foremost the impetus for beginning this study and for persevering through it comes from my advisor Dr. Chuang Wang. Second only to Dr. Wang are Dr. Lambert, Dr. Polly, and Dr. Bird who have all provided guidance and advice throughout this process. It cannot go unsaid that without the support of Dr. Polly, Dr. Wang, and Dr. Lambert through the MSP grant, I would not have had the resources to conduct this study or the dedicated time to finish. My current supervisor, Dr. David Wiley, has only known me a short time but has provided support and guidance for finishing this degree. Finally, the resources provided to me through the UNC Charlotte GASP program and the Department of Educational Leadership have been crucial to being able to begin and complete this degree and I am forever grateful for those opportunities.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

While there is an extensive body of literature on motivation and self-efficacy beliefs for students in mathematics (see Schunk, Zimmerman, and Pajares), less is understood about motivation and self-efficacy of mathematics teachers. Additionally, mathematics performance of students is measured regularly but teachers' mathematics content knowledge is seldom directly measured. A teacher's own mathematical abilities are only measured indirectly through an assumed relationship with their students' achievement. While there is a relationship between teacher's beliefs and practices, self-efficacy within a content area is often not considered. Therefore, measurement of a teacher's mathematics self-efficacy belief system may give insight into their mathematics teaching practice and subsequent student achievement in mathematics.

"Perceived self-efficacy is concerned with people's beliefs in their capabilities to produce given attainments" (Bandura, 2006, p.1). Self-efficacy is a future-oriented belief about the level of competence a person expects he or she will display in a given situation (Tschannen-Moran & Woolfolk Hoy, 2001). Self-efficacy in education is a construct that has been studied for more than 30 years. The study of self-efficacy began with the concept of teacher efficacy and has expanded to be applied to various contexts such as teacher job satisfaction, student academic achievement, and motivation (Tschannen Moran & Woolfolk Hoy, 2001). Self-efficacy has a central role in learning in that it contributes to one's motivation to learn (Bandura, 1977; Dornyei, 2009; Huang & Chang, 1998). Therefore, academic achievements are related to self-efficacy beliefs (Bong & Skaalvik, 2003; Huang & Chang, 1998). Researchers have also applied principles of self-efficacy to specific educational content areas such as literacy, science, and mathematics. The application of self-efficacy beliefs to specific tasks is more appropriate than the use of overarching measures of self-efficacy within a broad field and aligns more to Bandura's original thoughts about self-efficacy (Pajares, 1997).

In the literature, both terms "teacher self-efficacy" and "teacher efficacy" can be found and are used to represent the same concept. Bandura's social cognitive theory uses the terminology "self-efficacy" but he can also be cited using the word "efficacy" with the same meaning implied. Other researchers have used the term "efficacy" to mean the same as "self-efficacy." The term "self-efficacy" will be used most often in this research study unless specific references use the term "efficacy."

Teacher Self-Efficacy

The concept of teacher self-efficacy began with two questions on two Research and Development (RAND) Corporation evaluations of projects funded by the Elementary and Secondary Education Act (Armor et al., 1976; Berman, McLaughlin, Bass, Pauly, & Zellman, 1977; Tschannen Moran & Woolfolk Hoy, 2001). These two studies included two, 5-point Likert-type items ("When it comes right down to it, a teacher really can't do much because most of a student's motivation and performance depends on his or her home environment" and "If I try really hard, I can get through to even the most difficult or unmotivated students"). A total score was calculated for these items. It is unclear whether the RAND researchers intended to specifically measure teacher self-efficacy using those two items or whether the inclusion of these items was incidental. These two items sparked great interest in educational research because teacher efficacy had never been examined before. Attention became focused on the use of self-efficacy as a predictive factor related to student academic achievement (Tschannen Moran & Woolfolk Hoy, 2001).

The RAND studies used Rotter's Social Learning theory (Rotter, 1954) as a theoretical framework for self-efficacy when designing these items (Henson, 2001). Rotter's theory uses *locus of control* as an element of one's personality and thus their self-efficacy. His theory was well established in educational psychology literature long before Bandura's Social Cognitive Theory, which is now the most widely accepted framework for self-efficacy. Bandura's theory gained popularity in the late 1970's. Self-efficacy scales that began to emerge after the RAND study, such as Gibson and Dembo's (1984) Teacher Efficacy Scale (TES), claimed to align with Bandura's theory (Tschannen Moran & Woolfolk Hoy, 2001). However, Gibson and Dembo (1984) used the RAND items as a guide for creating the TES, and many subsequent self-efficacy scales used the TES as a model for creating new scales. This cross-contamination of self-efficacy theories left confusion in the field and called into question some of the established instruments for measuring teacher self-efficacy (Henson, 2001; Tschannen Moran & Woolfolk Hoy, 2001).

From the initial RAND studies, two branches of teacher self-efficacy emerged in the literature (Tschannen Moran & Woolfolk Hoy, 2001). Most modern educational psychologists prefer self-efficacy measures that align closely with Bandura's iteration of self-efficacy instead of Rotter's theory. Bandura's theory is comprised of two facets of self-efficacy: *outcome expectations* and *efficacy expectations*. Rotter's theory marries self-efficacy to a person's *locus of control* over a situation as a factor of their personality (Henson, 2001). Another stark contrast between Rotter and Bandura is that Rotter's theory does not include instincts or motivation and instead uses a person's environment as a major factor in their behavior while Bandura's theory relies heavily on motivation and instinct as a primary component of a person's behavior (Haines, McGrath, & Pirot, 1980; Henson, 2001).

In an attempt to better understand teacher self-efficacy, several researchers have created instruments to measure teacher self-efficacy. Specific information about the concept development of teacher self-efficacy will be discussed in Chapter 2. Although the history of the measurement of self-efficacy spans the last 30 years, presently the most widely-accepted measure of the concept is the *Teacher's Sense of Efficacy Scale*, or TSES (Tschannen Moran & Woolfolk Hoy, 2001). The TSES measures three constructs: *efficacy for instructional strategies, efficacy for student engagement,* and *efficacy for classroom management*. Tschannen Moran and Woolfolk Hoy (2001) hold closely to Bandura's original thoughts about self-efficacy and used his instrument, the Teacher Efficacy Scale (TES), as a model for the TSES. More information about this instrument and others will be discussed in Chapter 2.

Teacher Mathematics Self-Efficacy

While general teacher self-efficacy remains of interest to researchers, context-specific self-efficacy of both teachers and students has also become an area of interest to educational researchers. More recent efforts to establish benchmarks of teacher quality propose an obvious link between student achievement and teacher quality (Darling-Hammond & Youngs, 2002). Therefore, researchers have become interested in factors that can enhance student achievement, with one of those factors being teacher self-efficacy.

The prominence of results from international tests of academic achievement are also a motivating factor for this research since American student achievement in mathematics is much lower than student achievement in many European and Asian countries (National Center for Education Statistics [NCES], 2010). These test results, and other past international test results (NCES, 2007, 2009) have caused President Obama to declare this America's "Sputnik moment" where science and mathematics achievement again become the focus of American education (Lee, 2010).

Self-efficacy of students in mathematics has been studied to a great extent by several prominent researchers in the field and has been shown to be a motivator for increasing students' mathematics academic achievement (Pajares & Miller, 1994; 1995). It has also been shown to be a predictor of mathematics student achievement. High mathematics self-efficacy is related to better use of self-regulated learning strategies (Wang & Pape, 2007) and influences interest in and positive attitudes about mathematics (Bandura, 1997). It has also been shown to influence career choice behavior (Betz & Hackett, 1986).

Measuring Teachers' Mathematics Self-Efficacy

Researchers have attempted to measure teacher self-efficacy for teaching mathematics by creating their own instruments. Enochs, Smith, and Huinker (2000) adapted their Science Teaching Efficacy Beliefs Instrument (STEBI), a valid measure of science teaching self-efficacy, to be mathematics specific. It is called the Mathematics Teaching Efficacy Beliefs Instrument (MTEBI). The two constructs in the MTEBI are *personal mathematics teaching efficacy* and *mathematics teaching outcome expectancy*. Although the authors report that the MTEBI is both valid and reliable, they indicate that further work on validation of the instrument is needed. A recent study analyzing the MTEBI using Item Response Theory (IRT) provided evidence that the validity of the MTEBI was not as high as previously thought (Kieftenbeld, Natesan, & Eddy, 2010). Other smaller and lesser known instruments have been created for measuring mathematics self-efficacy but all have met the same challenges of not being a valid measure or failing to align with Bandura's theory.

Ernest (1989) touted that a mathematics teacher's belief system is complex and has three parts: ideas about mathematics as a subject for study, ideas about the nature of mathematics teaching and ideas about mathematics learning. One comprehensive way to measure teacher's beliefs about mathematics is to develop an instrument to measure self-efficacy for mathematics, and self-efficacy for teaching mathematics as these beliefs impact a teacher's practice.

The TSES, the only valid instrument for measuring general teacher self-efficacy, loses construct validity when its items are modified to be mathematics-specific (Swackhamer, 2010). Even modified to be mathematics-specific, it still lacks questions associated with sub-tasks specific to teaching mathematics (use of correct tools, differentiation of methodology, content, ideas about mathematics as a subject for study, ideas about the nature of mathematics teaching and ideas about mathematics learning). It is more appropriate to examine the subtasks associated with teaching mathematics and apply self-efficacy measures to those tasks, especially since overarching measures of self-efficacy are not the best measure of self-efficacy for completing specific tasks. However, too much specificity might hinder the ability to use self-efficacy as an indicator of performance (Pajares, 1997).

Another issue with current measures of teacher self-efficacy, including the TSES, is the lack of specificity to a teacher's grade level (elementary, middle, or high school). Educational experiences for teachers and students are different at each of these grade levels and thus expectations of a classroom teacher and sub-tasks associated with teaching content are different. Teaching academic content to students is a task that requires a specific skill set beyond classroom management and relationship building. Instruments created for the measurement of teacher self-efficacy in the past have not been grade-level specific, but for this study, elementary teachers are the focus.

Elementary teachers in the United States are often expected to be expertly teach all subjects to their students, but this may not be how they view their own abilities. Some elementary teachers have communicated their apprehension about mathematics and their own self-doubts regarding teaching mathematics (McGee, Wang, & Polly, in press; Piel & Green, 1993; Polly, McGee, Wang, Lambert, Pugalee, & Johnson, 2010). Because self-efficacy is context specific (Bandura, 1977), teachers often have varied beliefs toward different aspects of teaching mathematics. Some feel highly efficacious in their knowledge of content but less efficacious in their ability to teach those concepts to students. Others have content-specific self-efficacy beliefs and may feel extremely efficacious about teaching some aspects of the content (i.e. patterns) but less efficacious about teaching other content (i.e. fractions).

In an effort to address both the shortcomings of other mathematics self-efficacy instruments and to measure self-efficacy of elementary teachers specifically, this study considers the discrete nature of teaching elementary mathematics and specific aspects of that environment with regard to teachers' self-efficacy. Evaluating elementary teacher's self-efficacy for teaching mathematics could be a crucial step in understanding the performance of their students in mathematics. Understanding elements of an elementary teacher's mathematics self-efficacy beliefs and the relation of those elements to student learning outcomes could provide insight into ways to improve student learning in mathematics.

Theories about learning and motivation are important aspects of effective education, but theory has to be made applicable for practitioners. Thus, an instrument using Bandura's theory as a framework should include questions that address the overarching self-efficacy beliefs of mathematics teachers as well as their beliefs about completing specific sub-tasks associated with teaching mathematics.

Validating a Survey Instrument

The careful consideration of methodology in which to compose a measurement tool leads one to consider the survey as a viable template. Survey research is a necessary and valuable method of collecting data in educational research (Ebel, 1980). Information obtained through survey methodology is used to draw conclusions and can help educational leaders make decisions. In an age of accountability, survey research has the ability to answer some of education's most pressing questions.

A technically-adequate survey instrument must pass several rigorous tests to be considered useful for the purpose in which it was intended. Survey instruments must show evidence of validity, a property of the meaning of data obtained through a survey or test (Messick, 1995). Validity ensures that the interpretation of a survey outcome, or score, is accurate (Messick, 1995). It also ensures that the intended theoretical constructs are supported and that conclusions drawn from the survey are accurate and reliable (Colosi, 1997; Messick, 1995).

Validity is a global concept that is a collection of evidences, one of the most common being construct validity (Messick, 1995). Construct validity includes any evidence that is imperative to the interpretation of a test score, including the use of sub-scales and including content validity and criterion-related validity (Messick, 1995). Construct representation is one outcome of construct validity. This process relies heavily on cognitive psychology to determine how one's responses on a test or survey indicate task performance relevant to some underlying theoretical or psychological idea (Cronbach & Meehl, 1955; Messick, 1995). The theoretical orientation of the investigator is also a major factor in defining construct validity (Cronbach & Meehl, 1995). There are several other sub-measures of construct validity including:

- Concurrent validity, how well the survey correlates with other surveys that have been previously validated that measure similar constructs Gay, Mills, & Airasian, 2009),
- Convergent validity, the degree to which constructs on an instrument are related to similar constructs (Moss, 1992), and
- Discriminant validity, the degree to which constructs on an instrument are unrelated to dissimilar constructs (Moss, 1992).

Evidence of validity should be based on elements such as: test content, response processes, internal structure, relations to other variables, and the consequences of testing, according to *Standards for Educational and Psychological Testing* (AERA, 1999). All of these procedures provide evidence for the strength of the survey instrument.

The purpose of this study is to show evidence of reliability and validity for a survey instrument that measures the mathematics self-efficacy of elementary teachers. The following research questions were of interest:

1. What is the reliability of the newly-development instrument?
2. What is the structure of the newly-developed instrument?
3. What evidence exists regarding validity of the instrument?

This study focused on determining the structure of a newly-created instrument in an attempt to align that structure to Bandura's theories of self-efficacy. This study also focused primarily on two types of reliability: test/retest reliability and internal consistency. Test/retest reliability implies that each time a participant completes a survey or test they should score approximately the same. Internal consistency implies that the questions intended to measure the same concept do so on a consistent basis and in a consistent way (Colosi, 1997). Subsequently, content validity and construct validity, specifically concurrent and criterion-related validity were of interest in this study.

Context for the Study

Data for this study were collected as a part of the external evaluation of a Mathematics Science Partnership (MSP) grant project. This project was funded by a three year, $2.4 million dollar grant from the United States Department of Education, funded through the North Carolina Department of Public Instruction. The MSP grant project is a collaborative partnership between two school districts and the university where the researcher is a student. The purpose of the MSP grant project is to develop standards-based elementary mathematics educators by providing intensive professional development and support centered on the *Investigations in Number, Data, and Space* (Investigations) curriculum. The MSP grant project took place over a three year period from August 2009 to August 2012. Three cohorts of elementary teachers participated in the MSP grant project during this time. Data collected from those three cohorts are used in this study.

This dissertation study took place in four phases. Phases I and II used data collected from Cohort II (2010-2011) of the MSP grant project and Phases III and IV used data collected from Cohort III (2011-2012) of the MSP grant project. In Phase I, the participants answered questions on the newly-developed instrument as a pre-test in August 2010. Data from this phase of data collection was used to determine reliability of the new instrument by determining the correlation coefficient of the items within each construct. This data was also used to explore the structure of the constructs in the instrument using Exploratory Factor Analysis (EFA) and to explore concurrent validity by correlating the new instrument to the TSES.

In Phase II of this study, participants responded to items on the same, unaltered instrument as a post-test measure in February 2011. These data were used to examine correlations between pre and post-test administrations of the SETMI. Each participant's pre-test score will be compared to their post-test score on the new instrument using a correlation coefficient.

Phases III and IV of this study took place during year three of the MSP grant (2011-2012). In June 2011, the researcher met with an elementary mathematics expert from the university to revise mathematics content items in order that they aligned better to the state standard course of study. Items were also revised to clarify the content being assessed and to shorten the survey length if possible. The basis for this change was also influenced by output from the EFA. Afterwards, a focus group of elementary teacher-leaders from Cohort II of the MSP grant met in July 2011 to discuss the new items on the instrument. Five elementary teachers provided feedback to the researcher regarding item wording. Only one item was suggested to be revised. This revision to the instrument was made and the Cohort III participants responded to items on the instrument as a pre-test measure for MSP grant purposes in August 2011 (Phase IV). This set of data was used to conduct a Confirmatory Factor Analysis (CFA).

Delimitations

There are a few delimitations, or limiting factors, for this study. Participants in this study are demographically representative of the population of teachers in both districts. In fact, in District Two, nearly all K-5 teachers in the district will have completed the MSP grant project by the end of year three. However, personal characteristics of the teachers who are participants in this study may differ greatly from the general population of teachers for several reasons. First, MSP grant participants are selected through an application process that is managed by each partner school district independently. Teachers are selected based on criteria set by the school district. This may lead to differences in the participants as a sample group, specifically about mathematics or the curriculum. Second, teachers who participated in the MSP grant project alongside their school colleagues had a different experience with the project than teachers who were the sole participants from their schools or grade levels, which might influence their self-efficacy. Another limiting factor of this study is the acute focus on elementary mathematics teacher self-efficacy. Differences between this subgroup of teaching professionals and other groups assessed in previous self-efficacy research may lead to different conclusions about self-efficacy.

Limitations

The limitations of this study include context-specific elements that may influence its outcomes. Participants of this study, by the nature of their participation in the MSP grant, completed the newly-developed instrument as a part of their requirements of the grant. Although these participants are volunteers, most complete the survey instrument developed for this study in concert with a packet of other instruments. Participants completed three surveys in one administration, of which only one is used in this study. Their responses to the other two surveys might have influenced their choices on the self-efficacy instrument. Additionally, completing several surveys at once could lead the participant to rush to finish all surveys without taking the proper time to thoughtfully respond to the items. Another possible limitation could be due to the nature of self-report responses that are susceptible to the effects of social desirability bias (Edwards, 1957).

Second, participants in this study are compensated as a part of their participation in the MSP grant project and therefore compensated in part for completion of the survey instrument developed for this study. By receiving compensation to complete the surveys, participants might focus more on the completion of the survey in a short period of time instead of thoughtful responses to the items. Surveys are often given to participants at the end of the professional development program day, thus potentially amplifying this affect.

Third, participants in Phases I and II of this study, by nature of their participation in the MSP grant project may already have a high self-efficacy for teaching mathematics or have indicated that they would like to increase their mathematics teaching self-efficacy by participation in the MSP grant project. Teachers are selected for participation in the MSP grant based on criteria set by each school district independently. The goal of District Two was to have most elementary teachers complete the MSP grant program by the end of year three. The goals of District One, a much larger district, are varied. Selection of District One teachers for participation in the grant is determined by factors such as: how many teachers from their school applied to the grant, the history of the school as participating in the MSP grant project, the cooperation of the school building principal, and the availability of the teacher to attend the summer workshop. Data from this group of teachers potentially may not generalize to all elementary school teachers because of either their already-existing level of self-efficacy for teaching mathematics, or other individual factors.

Assumptions

The primary assumption associated with this study is that social desirability does not impact the responses of participants (Edwards, 1957). This could be a concern because data collection for this study is part of an evaluation of a professional development program. Additionally, data collection in this study hinges on self-report. Self-report is a widely-used data collection method for studies concerning psychological concepts such as self-efficacy. It holds much value for behavioral and social research (Baldwin, 2009; Harrell, 1985).

Definitions

For the purposes of this research study, several variables are operationally defined as follows:

- Self-efficacy: beliefs about how well one can perform a specific task with the assessment of one's ability (Bandura, 1994).
- Self-efficacy for mathematics: beliefs about how well once can perform a specific mathematical task with the assessment of one's ability (Kahle, 2008).
- Self-efficacy for teaching mathematics: beliefs about how well one can perform specific duties relevant to teaching mathematics to another with the assessment of one's ability (Tschannen-Moran & Woolfolk Hoy, 2001). In this study, although self-efficacy for mathematics differs from self-efficacy for teaching mathematics in the literature, the researcher believes that an important element of being efficacious at teaching mathematics is being efficacious in performing mathematical tasks. Therefore, both elements are included in the review of literature and on the SETMI itself.
- Mathematics Science Partnership [MSP]: A grant from the United States Department of Public Instruction given through state department of education offices in an effort to increase mathematics or science student achievement. The context of this study falls within the context of an MSP grant project of which the researcher is an external evaluator.
- Reliability: the ability of a test or survey to remain stable over time. This is sometimes referred to as internal consistency (Schumacker & Lomax, 2010).
- Test/retest reliability: an indication of the extent to which participants' scores on an assessment are similar on repeated assessments (Colosi, 1997).
- Validity: an overall evaluative judgment concerning the interpretation of test scores and is not a property of the test itself (Messick, 1995).

- Construct validity: the extent to which scores on a measure accurately represent a theoretical construct (Schumaker & Lomax, 2010). Validity is an evaluation of the hypothesized relationships between theoretical constructs and their observed counterparts (Cronbach & Meehl, 1955).
- Content validity: establishing that test items are well representative of the universal knowledge within a specific area (Cronbach & Meehl, 1955).
- Concurrent validity: an evaluation of how well a new measure correlates with a test of similar content that has previously been validated. This is particularly important when one instrument is proposed as a substitute for another (Cronbach & Meehl, 1955).

Chapter Summary

The purpose of this study was to develop and validate a survey instrument intended to measure self-efficacy beliefs of elementary mathematics teachers. Self-efficacy is a belief about one's ability to perform a specific task. With regards to teaching in general, several self-efficacy constructs have been examined in the literature, most aligning to Bandura's theory of social cognition and his thoughts about self-efficacy. These include efficacy *for instructional strategies* and *efficacy for student engagement*, both a part of a widely-used instrument intended to measure general teaching self-efficacy entitled the TSES. Bandura (2006), as well as other researchers, have touted that a more valid way of measuring self-efficacy is to measure self-efficacy within the context of performing specific tasks. Other researchers have argued that too much specificity may hinder the ability to use self-efficacy as an indicator for performance. However, a medium between both ways of thinking is possible if caution is undertaken in producing a survey instrument.

In order to measure self-efficacy specific to mathematics teaching in elementary school, elements from the TSES were combined with content questions to create the Self-efficacy for Teaching Mathematics Instrument (SETMI). This instrument was created in July 2010. The purpose of this study was to validate the survey instrument by examining its structure and fit to the theoretical model.

CHAPTER 2: REVIEW OF LITERATURE

This chapter will explore the facets of self-efficacy research that abound within the literature. Self-efficacy, as introduced by teacher self-efficacy research and Bandura's Social Cognitive Theory will be discussed. Furthermore, the measurement of teacher self-efficacy is explained, both generally and within contexts. A detailed explanation of how the most widely used instrument to measure teacher self-efficacy, the TSES, was created and validated is also included in this chapter. Finally, this chapter will discuss the need for a new instrument to measure elementary teacher's mathematics self-efficacy and how evidence of validity can be obtained and measured.

Self-Efficacy

Self-efficacy stems from Social Cognitive Theory, a learning theory consolidated by Bandura beginning in the 1970's which is best represented by a triadic model that describes the relationships between one's behaviors, the environment, and personal factors (Bandura,1997; Pajares, 2002). Social Cognitive Theory is a re-conceptualization of Rotter's (1954) Social Learning Theory. Both Bandura and Rotter were interested in expectancy, although they define expectancy quite differently (Haines, McGrath, & Pirot, 1980). Rotter's work uses *locus of control of reinforcement* (*locus of control*) as a generalized expectancy and thus a dimension of one's personality. Rotter (1954) dictates that actions are determined by reinforcement which comes from either internal or external feelings of control and that these feelings are over-arching, generalized beliefs that guide one through daily life. Bandura (1977) contends that personal competence for completing a specific task is the most influential factor in determining one's beliefs about behaviors and thus one's ability to produce desired action. According to Bandura, "Perceived self-efficacy is concerned with people's beliefs in their capabilities to produce given attainments" (2006, p.1).

Self-efficacy and locus of control are also sometimes incorrectly viewed as measurements of the same phenomenon at different levels of specificity (Bandura, 1997). "Beliefs about whether one can produce certain actions [self-efficacy] cannot, by any stretch of the imagination, be considered the same as beliefs about whether actions affect outcomes [locus of control]" (p. 20) as these concepts bear little or no relationship to each other empirically (Bandura, 1997). Therefore, *self-efficacy* beliefs are task-specific and alterable as opposed to *locus of control of reinforcement* which is generalizable and constant (Haines et al., 1980).

Self-efficacy is also commonly confused with self-concept and self-esteem (Bandura, 2006; Pajares, 1997). "Self-concept is a composite view of oneself that is presumed to be formed through direct experience and evaluations adopted from significant others" (Bandura, 1997, p. 10). Self-concept can be measured by asking people to rate how well different attributes apply to themselves. The effect of self-concept on self-efficacy is weak and one's personal attributes may or may not be relevant to their efficacy for completing a task or producing an outcome (Bandura, 1997). Self-esteem is a judgment of one's self-worth and is entirely different from self-efficacy, or judgment of one’s capability (Bandura, 1997). Bandura stated, “People need much more than high self-esteem to do well in given pursuits” (1997, p. 11).

Another conflicting concept with self-efficacy is the notion of *outcome expectancies*. *Outcome expectancies* are judgments of an outcome of behavior (Pajares, 2002). Rotter’s locus of control provides one example of *outcome expectancies*. Beliefs about outcome causality need to be conceptually separated from self-efficacy as “performance is causally prior to outcomes” (Bandura, 1997, p.21). However, perceived outcomes do factor into self-efficacy beliefs (Pajares, 2002). One who is efficacious in their performance, and receives environmental feedback from those performances will come to expect positive outcomes (Bandura, 1997). Similarly, one who is efficacious but who does not receive positive environmental feedback from performance will come to expect different outcomes. As self-efficacy drives performance, the two concepts are related but do not describe the same phenomenon (Bandura, 1997).

In addition to influencing performance, self-efficacy influences the choices people make, how much effort they will expend on an activity, how long they will persevere, and how resilient they will be in the face of adversity (Pajares, 1997). It is important because unless people believe their actions can produce the outcomes they desire, they will not be motivated to overcome obstacles (Pajares, 2002). Self-efficacy is also seen as a future-oriented belief about the level of competence a person expects he or she will display in a given situation (Tschannen-Moran & Woolfolk Hoy, 2001). A strong sense of efficacy can be motivating and sustaining while a weak sense of efficacy can be debilitating and lead one to shy away from difficult tasks (Pajares, 1997). A strong sense of self-efficacy also influences the amount of stress and anxiety one may feel when attempting to accomplish seemingly difficult tasks (Bandura, 1994; Pajares, 1997; Zeldin & Pajares, 2000).

Self-efficacy has a central role in learning in that it contributes to motivation (Bandura, 1977; Dornyei, 2009; Huang & Chang, 1998). Therefore, academic achievements are affected by self-efficacy beliefs (Bong & Skaalvik, 2003; Huang & Chang, 1998). Self-efficacy is mediated by executive processes like planning and goal setting (Bandura, 1994). By imaging future outcomes, people can generate motivators for their behavior. If a person becomes motivated to learn by future prospects while also feeling efficacious in their pursuits, success is more likely and persistence towards the goal is more purposeful (Bandura, 1977; Vick & Packard, 2008).

As integral as self-efficacy may be in many cognitive processes and subsequent behaviors, it is not an overarching measure of self-ability for each individual (Bandura, 1977). There can be no global measure of self-efficacy (Bandura, 2006). Judgments of self-efficacy are task specific and vary in strength and magnitude (Bandura, 1977; Bong & Skaalvik, 2003; Pajares, 1997; Wang & Pape, 2007). Self-efficacy is also not solely responsible for the outcome of an event but the outcomes that one may expect are dependent on one’s judgments of how much they can accomplish (Bandura, 1984).

Self-efficacy beliefs are complex. Often, they are better predictors of performance than actual ability, previous attainments, knowledge, or skill (Pajares, 2002). Highly efficacious people may not always act in ways consistent with those beliefs if the outcome of their actions contradicts their desired outcome. People with lower efficacy may use positive outcomes to enhance their efficacy in a certain area (Pajares, 2002).

Self-efficacy has persisted as a fundamental concept for partially explaining actual versus expected outcomes of behavior.

Teacher Self-Efficacy

Self-efficacy research in education began with the study of teacher self-efficacy (Tschannen Moran & Woolfolk Hoy, 2001). Although the study of self-efficacy in education began with teachers, the focus of self-efficacy and motivation research has largely been on students for the last couple of decades. Self-efficacy has become an important factor in discovering components of motivation for learning in students (Schunk, 1991). In the current educational climate of accountability initiatives, self-efficacy beliefs of the teacher may be an important link in discovering differences between actual versus expected outcomes.

There is notably some coherence between a teacher's beliefs and their practices in the classroom (Peterson, Fennema, Carpenter, & Loef, 1989; Stipek, Givvin, Salmon, & MacGyvers, 2001; Williams & Williams, 2010). Bandura (1994) asserts that teachers who have a high self-efficacy about their teaching can motivate students to learn. He also states that those with low instructional self-efficacy may use more teacher-centered methods of teaching and motivate using mostly negative reinforcement. Teacher self-efficacy is defined by Woolfolk and Hoy (1990) as a teacher's belief that he or she can influence student learning. Dellinger, Bobbett, Olivier, and Ellett (2008) define teacher self-efficacy as a "teacher's individual beliefs in their capabilities to perform specific teaching tasks at a specified level of quality in a specific situation."

Dellinger et al. (2008) disagree somewhat with the Woolfolk and Hoy (1990) definition of teacher self-efficacy and assert that teacher "self-efficacy" is actually teacher “efficacy.” Dellinger et al. (2008) claim that “efficacy” is focused on an outcome irrespective of teacher’s beliefs about their ability to perform certain teaching tasks required in specific contexts (Dellinger et al., 2008). Most literature does not make a distinction between the two terms (Dellinger et al., 2008). In fact, there are several contradictory definitions for teacher self-efficacy leading to confusion within the field about how to accurately measure the construct (Pajares, 1997; Tschannen-Moran & Woolfolk Hoy, 2001). For the purposes of this study, the term teacher “self-efficacy” will be used primarily to indicate that it is an internally initiated belief which is linked to Bandura’s Social Cognitive Theory, the most widely accepted theory of self-efficacy to date.

Measuring Teacher Self-efficacy

The measurement of teacher self-efficacy has a history of more than 30 years. The terms “teacher efficacy” were first used in two reports of RAND Corporation evaluations of projects funded by the Elementary and Secondary Education Act (Armor, et al., 1976; Berman, McLaughlin, Bass, Pauly, & Zellman, 1977). These two studies evaluated the concept of teacher efficacy by computing a total score for responses to two, 5-point Likert-type items: *(a) When it comes right down to it, a teacher really can’t do much because most of a student’s motivation and performance depends on his or her home environment* and *(b) If I try really hard, I can get through to even the most difficult or unmotivated students* (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998; Woolfolk & Hoy, 1990). The RAND study used Rotter’s Social Learning Theory (1954), as a theoretical basis, which cites locus of control as a component of efficacy (Woolfolk & Hoy, 1990). Bandura’s Social Cognitive Theory had not yet been published.

The RAND study spurred a discussion in research literature about teacher self-efficacy, how to measure the construct, and what other factors are related. Self-efficacy was an elusive concept that suddenly captured the imagination of researchers because it appeared integral to a teacher's perceived and actual effectiveness (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998). The concept of teacher efficacy has been measured in numerous ways in earlier literature (See Table 1 for a listing of instruments).

Table 1.

Complete Listing of Self-Efficacy Instruments

Author	Title	Theoretical Basis	Sample Items
Armor et al. (1976) Berman et al. (1977)	RAND Study	Rotter's Social Learning Theory (*locus of control of reinforcement)*	If I try really hard, I can get through to even the most difficult or unmotivated students.
Rose & Medway (1981)	Teacher Locus of Control questionnaire (TLC)	Rotter's Social Learning Theory (*locus of control of reinforcement)*	When the grades if your students improve, it is more likely: a. because you found ways to motivate the students, or b. because your students were trying harder to do well. c.
Guskey (1981)	Responsibility for Student Achievement questionnaire (RSA)	Rotter's Social Learning Theory (*locus of control of reinforcement)*	If a student does well in your class, would it probably be: d. Because that student had the natural ability to do well, or e. Because of the encouragement you offered.

(continued)

Table 1.

Complete Listing of Self-Efficacy Instruments

Author	Title	Theoretical Basis	Sample Items
Ashton et al. (1982)	Webb Efficacy Scale	Rotter's Social Learning Theory (*locus of control of reinforcement)*	a. A teacher should not be expected to reach every child; some students are not going to make academic progress. b. Every child is reachable. It is a teacher's obligation to see to it that every child makes academic progress. Circle One: 1. I agree mostly with A. 2. I agree mostly with B.
Ashton et al. (1982)	Ashton Efficacy Vignettes	Bandura's theory of self-efficacy from Social Cognitive Theory	One of your students misbehaves frequently in your class and is often disruptive and hostile. Today in class he began roughhousing with a friend in the back of the class. You tell him firmly to take his seat and quiet down. He turns away from you, says something in a belligerent tone that you can't hear and swaggers to his seat. The class laughs and then looks to see what you are going to do. How effective would you be in responding to this student in a way that would win the respect of the class? (Likert-scale from 1 "Extremely Ineffective" to 7 "Extremely Effective")

(continued)

Table 1.

Complete Listing of Self-Efficacy Instruments

Author	Title	Theoretical Basis	Sample Items
Betz & Hackett (1983)	Mathematics Self-Efficacy Scale (MSES)	Bandura's theory of self-efficacy from Social Cognitive Theory	Add two large numbers (e.g., 5739 + 62543) in your head. (Likert-scale from 0 "No confidence at all" to 9 "complete confidence).
Gibson & Dembo (1984)	Teacher Efficacy Scale (TES)	Bandura's theory of self-efficacy from Social Cognitive Theory	When a student does better than usually, many times it is because I exert a little extra effort. (Likert-scale from 1 "Strongly Agree" to 6 "Strongly Disagree")
Riggs & Enochs (1990)	Science Teaching Efficacy Belief Instrument (STEBI)	Bandura's theory of self-efficacy from Social Cognitive Theory with items taken from the TES	When a student does better than usual in science, it is often because the teacher exerted a little extra effort. (Likert-scale from "Strongly Agree" to "Strongly Disagree")
Bandura (1997)	Teacher Self-efficacy Scale (TSS)	Bandura's theory of self-efficacy from Social Cognitive Theory	How much can you do to keep students on task on difficult assignments? (Likert-scale from 1 "Nothing" to 9 "A Great Deal")
Kranzler & Pajares (1997)	MSES-Revised (MSES-R)	Bandura's theory of self-efficacy from Social Cognitive Theory	Same as the MSES with a Likert Scale from 1 to 5.

(continued)

Table 1.

Complete Listing of Self-Efficacy Instruments

Author	Title	Theoretical Basis	Sample Items
Enochs et al. (2000)	Mathematics Teaching Efficacy Belief Instrument (MTEBI)	The STEBI modified to be math-specific. Bandura's theory of self-efficacy from Social Cognitive Theory with items taken from the TES	When a student does better than usual in mathematics, it is often because the teacher exerted a little extra effort. (Likert-scale from "Strongly Agree" to "Strongly Disagree")
Tschannen-Moran & Woolfolk Hoy (2001)	Teacher's Sense of Efficacy Scale (TSES) formerly the Ohio State Teacher Efficacy Scale	Bandura's theory of self-efficacy from Social Cognitive Theory with Likert-scale from the TSS	How much can you do to get through to the most difficult students? (Likert-scale from 1 "Nothing" to 9 "A Great Deal")
Dellinger et al. (2008)	TEBS-Self	Bandura's theory of self-efficacy from Social Cognitive Theory	Right now in my present teaching situation, the strength of my personal beliefs in my capabilities to… 1. Plan activities that accommodate the range of individual differences among my students… (Likert-scale from 1 "weak beliefs in my capabilities" to 4 "very strong beliefs in my capabilities).

Three measurement instruments emerged in the literature that used Rotter's (1954) concepts of *internal-external locus of control of reinforcement* as a theoretical framework: the Teacher Locus of Control questionnaire (TLC) (Rose & Medway, 1981), the Responsibility for Student Achievement questionnaire (RSA) (Guskey, 1981), and the Webb Efficacy Scale (Ashton et al, 1982) (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998). All three of these instruments were never adopted and have all but disappeared from the literature after their initial conception (Tschannen-Moran & Woolfolk Hoy, 2001). It is difficult to assume exactly why this is the case, but an analysis of the items show evidence of inconsistency in theoretical underpinnings of the instrument and large similarities between the instruments. Bandura's Social Cognitive Theory also emerged just before these instruments were published, which may have interfered with their popularity and use.

Bandura's Social Cognitive Theory (1977) consists of two main constructs: *efficacy expectations* and *outcome expectations*. E*fficacy expectations* are an individual's convictions for orchestrating the necessary actions to complete a task. *Outcome expectations* are defined as an individual's expectations for performing a task that will lead to a certain outcome. Current use of the term *self-efficacy* comes from the original construct of *efficacy expectations*. The distinction between both constructs is important as *efficacy expectations* is the link between Bandura's theory and the current work in self-efficacy. Bandura (1986) also touted that *outcome expectations* add little to the predictive power of efficacy measures, but can provide incentives. Other work on the topic has diluted the operational definition of self-efficacy by including elements in measurement of self-efficacy that are more relevant to *outcome expectations* (Pajares, 1997).

The emergence of Bandura’s (1977) theory caused work within self-efficacy to diverge into alignment with either Bandura’s theory or Rotter’s theory. One early self-efficacy instrument by Gibson and Dembo (1984) took the two items from the RAND studies and attempted to expand them, while also combining them with two elements of Bandura’s newly published theory: outcome and efficacy expectations (Tschannen Moran & Woolfolk Hoy, 2001). The two constructs were renamed *general teaching efficacy* and *personal teaching efficacy* and are the theoretical basis for the Teacher Efficacy Scale (TES) (Gibson & Dembo, 1984). One sample item from that scale is, “When a student does better than usually, many times it is because I exert a little extra effort. (Likert-scale from 1 “Strongly Agree” to 6 “Strongly Disagree”).” Although the TES is a foundational work used to create other self-efficacy measures, items like the one above appear to align more with Rotter’s *locus of control* than with Bandura’s *self-efficacy* as they elude that the teacher’s actions influences the student’s behavior instead of measuring the teacher’s own thoughts about their ability to produce a behavior. It is important to recall that the RAND studies used Rotter’s theory as a theoretical framework (Henson, 2001).

Other researchers used the Gibson and Dembo (1984) study as a springboard to measure self-efficacy within specific contexts by rewording the TES items to be content specific (see the Science Teaching Efficacy Belief Instrument [STEBI] by Riggs & Enochs, 1990) or by creating hypothetical context-specific situations for teachers to respond to (see the Ashton Vignettes by Ashton, Burr, and Crocker, 1984) (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998). Still, other researchers combined items from several of the aforementioned instruments in an effort to measure teacher self-efficacy (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998). None of the more brief instruments created in this manner have been used extensively in the literature. This might be due to the failure of researchers across the board to agree on one factor structure in the measurement of self-efficacy.

In 1990, Woolfolk and Hoy conducted an analysis of the concepts of teacher efficacy in the Ashton and Webb (1986) and Gibson and Dembo (1984) studies compared with Bandura's Social Cognitive Theory. They determined that the two constructs of *teaching efficacy* and *personal teaching efficacy* were actually not congruent with Bandura's constructs of *outcome expectations* and *expectancy expectations*. Henson (2001) has also published work on this topic.

In response to their findings, Woolfolk and Hoy (1990) then sought to explore perspective teacher's sense of efficacy. They proposed that teacher's efficacy was comprised of two different un-related components: *teaching efficacy* and *personal efficacy*. Further studies of these two factors have found them to be weakly related with correlation coefficients ranging from .15 to .20. In addition, further studies of both pre-service and in-service teachers have found that from 18% to 30% of the variance between teachers is explained by these two factors (Tschannen-Moran & Woolfolk Hoy, 2001). Woolfolk & Hoy (1990) warned against using a composite score for efficacy without taking these two attitudinal aspects into consideration.

The creation of so many diverse measurement tools for teacher self-efficacy created confusion about the nature of self-efficacy itself (Henson, 2001; Pajares, 1997; Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998). In response to the confusion, Bandura (1997) created his own Teacher Self-efficacy Scale (TSS). He further outlined that *locus of control* and *self-efficacy* are not empirically related and that *locus of control* is a weak predictor of behavior, denouncing Rotter's Social Learning Theory as a basis for teacher self-efficacy. Bandura's TSS contains seven constructs: (a) efficacy to influence decision making, (b) efficacy to influence school resources, (c) instructional efficacy, (d) disciplinary efficacy, (e) efficacy to enlist parental involvement, (f) efficacy to enlist community involvement, and (g) efficacy to create a positive school environment. The number of constructs included in the TSS is an attempt to capture more context specific information about teacher self-efficacy, as Bandura found most instruments overly generalized and poor measures of the specific tasks teachers are required to perform (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998). The instrument measures the seven constructs along a 9-point Likert scale, congruent with Bandura's (2006) view that response scales should range from either 0 to 100 or 0 to 10 in order to provide enough variance because people usually avoid extreme positions (Tschannen-Moran & Woolfolk Hoy, 2001).

Although the TSS was created by the father of self-efficacy theory, reliability and validity information for this instrument is not available because it was never published (Tschannen-Moran & Woolfolk Hoy, 2001). Tschannen-Moran and Woolfolk Hoy (2001) therefore found it necessary to create a better measure for teacher self-efficacy. They conducted a seminar entitled S*elf-efficacy for Teaching and Learning* at The Ohio State University. The seminar consisted of two researchers and eight graduate students, four of whom were practicing teachers. Seminar participants were charged with creating a new self-efficacy scale and chose to model their design on Bandura's TSS, although they did not agree with his seven constructs.

Creating the Ohio State Teacher Efficacy Scale (OSTES). Seminar participants created items for the instrument and titled it the Ohio State Teacher Efficacy Scale (OSTES). The OSTES was tested for validity in a series of three studies. The first study's purpose was to test the factor structure of the items that were created by seminar participants and determine which items to examine. The second study was conducted to determine the factor structure of the revised items. The third study was conducted to reinforce the factor structure of the revised measure. Information about each of these studies is discussed below and comes from the work on this topic by Tschannen-Moran and Woolfolk Hoy (2001).

Study #1 had 224 participants, 146 of which were pre-service teachers. The purpose of this study was to determine the factor structure of the items that were created by seminar participants. After conducting a factor analysis, 10 factors emerged with eigenvalues over one, accounting for 57.2% of the variance. The unrotated factor matrix was examined after the rotation failed to converge after 25 iterations. Only the first factor was examined, since it contained most of the variance (39.9%). Items with loadings higher than .60 were examined, which yielded 31 items with loadings ranging from .62 to .78. In response to this analysis, 32 of the original 52 items were maintained and further tested. Results of this study showed that some of the items designed to be accurate measures of self-efficacy were not. The loading of most items into one factor indicates that more diversity in item wording was needed to accurately measure different components of self-efficacy, but that *self-efficacy* is a second level factor, underlying even a simplified factor construct.

Study #2 had 217 participants, 70 of which were pre-service teachers. The participants were students at three different universities. It is possible that the researchers also needed a larger sample size to accurately examine the factor structure. Eight factors emerged from the factor analysis with eigenvalues greater than one, accounting for 63% of the variance. Using a scree plot, either two or three factors could be extracted. After examining both factor structures, a three-factor structure was chosen. Eighteen of the thirty-two items were selected for further testing. Items with low factors loadings were removed. The 18 items were then tested again and three factors emerged, accounting for 51% of the variance. Upon examining the 18 items (i.e., "How much can you do to get through to the most difficult students?", "How well can you respond to difficult questions from your students?"), it was concluded that the three factors could be labeled: *efficacy for student engagement* (8 items), *efficacy for instructional strategies* (7 items), and *efficacy for classroom management* (3 items). Although *efficacy for classroom management* was not one of the original intended constructs of teacher self-efficacy, the researchers and their seminar participants agreed that classroom management was an integral part of teaching and should be included. The researchers also determined that a subscale score could be computed for each construct, giving teachers a score for each aspect of teacher self-efficacy. These subscale scores were computed by calculating the mean of the responses to the items in each subscale.

The researchers chose to further test the three factor model by combining the responses from Study #1 and Study #2. This gave an overall sample size of 441. After conducting a factor analysis, one factor emerged that had factor loadings of .74 to .84. Because the three subscales were positively correlated, there was evidence for a second-order factor that could be called the *underlying construct of self-efficacy*. Because of the item loadings into one factor at this point, the researchers concluded that it would be appropriate to use the instrument to provide an overall score of teacher self-efficacy along with the three sub-scores.

Construct validity of the 18-item measure was tested by correlating the measure with the following existing measures: RAND items, the Hoy and Woolfolk 10-item adaption of the Gibson and Dembo TES in 1993, the pupil control ideology form from Willower, Eidel and Hoy in 1967, and the work alienation scale from Forsyth and Hoy in 1978). Total scores on the OSTES were positively correlated to the two RAND items (r = .35 and .28, $p < .01$) and to the *personal teaching efficacy* (r = .48, p < .01) and *general teaching efficacy* (r = .30, p < .01) factors of the Gibson and Dembo measure. Both the pupil control ideology form and the work alienation scale were found to be negatively correlated with the 18-item measure. The work alienation scale was used to measure discriminant validity as it was perceived to be negatively correlated with self-efficacy.

Following the thorough examination of the three factor model with 18 items, the researchers concluded that a third study (Study #3) was needed in order to bolster the weakness of the *classroom management* construct in relation to the other two constructs. Because the *classroom management* construct only contained 3 items, more were added using Emmer's 1990 teacher efficacy for classroom management scale as a guide. The researchers concluded that *classroom management* oriented survey items on most previous instruments had been worded towards managing difficult students, thus the new items were worded to include management of capable students.

The new instrument, with the additional questions in each construct, contained 36 items. The instrument was field tested in an educational psychology class at The Ohio State University consisting of 19 students. These students examined the content reliability of the items. The instrument was then tested with a sample of pre-service (n = 103) and in-service (n = 255) teachers totaling 358. Participants came from three different universities. A factor analysis was conducted and four factors emerged, accounting for 58% of the variance. One factor was considered to be the *underlying construct of self-efficacy,* and the other three considered: *efficacy for student engagement*, efficacy *for instructional strategies*, and *efficacy for classroom management.* Overall, the researchers documented the factor structure they expected but, due to some items having low factor loadings, the instrument was reduced to 24 items, leaving 8 items in each construct. The now 24-item structure was tested for reliability and found satisfactory.

Since the instrument was still quite long, and the factor loadings of some of the questions were very high, four of the eight items in each sub-scale with the highest factor loadings were selected to be tested for their reliabilities. This now shortened version of the instrument gave rise to both a Long Form and a Short Form of the OSTES. Further tests on both the Long and Short Forms of the OSTES revealed a four factor structure as well (accounting for 75% and 68% of the variance respectively). This finding validated earlier results and indicated that the *underlying construct of self-efficacy* could be measured by giving a total score, and the three sub-scales could be given scores as well. In evaluating pre-service versus in-service teachers, since both were included in the participant groups of each study, the researchers determined that a total score was more appropriate for pre-service teachers and total and sub-scale scores appropriate for in-service teachers. As pre-service teachers have not begun teaching, sub-scale scores bear little relevance.

The construct validity of the Long and Short Forms was tested by correlating the total scores on the OSTES responses of participants in Study 3 to: the RAND items and the Hoy and Woolfolk 10-item adaption of the Gibson and Dembo TES in 1993. Correlations were moderate for the RAND items ($r = .18$ and $.53$, $p < .01$) and the *personal teaching efficacy* ($r = .64$, $p < .01$) and *general teaching efficacy* ($r = .16$, $p < .01$) factors of the Gibson and Dembo measure. The researchers concluded from these correlations that the *personal teaching efficacy* construct captures the essence of efficacy better than the *general teaching efficacy* construct and that more work is needed on validation of the OSTES, although the factor structure has remained stable (Tschannen-Moran & Woolfolk Hoy, 2001). It is important also to note that in later studies, the OSTES is renamed the Teacher Sense of Efficacy Scale (TSES) and will be referred to as such throughout the conclusion of this dissertation.

Newer measures of teacher self-efficacy. Both the TES and the TSES are used most often to measure teacher efficacy in international and American research on the topic (Dellinger et al., 2008; Swackhamer, 2010) but this has not prevented other researchers from attempting to create other instruments. Dellinger et al. (2008) created the Teachers' Efficacy Beliefs System – Self Form (TEBS-Self) in an attempt to distinguish between concepts of *teacher efficacy* and *teacher self-efficacy*. The researchers make a compelling argument for the differences between teacher efficacy and teacher self-efficacy, but the TEBS-Self items themselves are not conceptually different from those in the TSES, even though they do address specific contexts. The TEBS-Self has not been widely used in the literature, other than in some dissertation studies from the researchers' institution (see Nolan, 2009).

The TSES has perhaps undergone the most rigorous tests of validity and reliability than other instruments introduced in the field (Swackhamer, 2010). However, the TSES items were intended for use with in-service teachers, but the validation of the factor structure was only performed using data from pre-service teachers. This could create some conceptual issues. Another issue is that although the TSES is considered a well-constructed instrument to measure general teacher self-efficacy, researchers agree with Bandura that more specific measures of self-efficacy are better indicators of a teacher's true self-efficacy for teaching (Henson, 2001; Pajares, 1997). The tasks associated with teaching are relative to the content being taught. Therefore, examination of teacher self-efficacy within context is the next logical step for this field of research.

Mathematics Beliefs of Teachers

Mathematics has become a subject of much discussion in the past decade. The No Child Left Behind (NCLB) legislation influenced curriculum by highlighting the importance of student achievement in reading and mathematics above other subjects. Additionally, efforts to establish benchmarks for "highly qualified teachers" in alignment with NCLB propose an obvious link between student learning and teacher quality (Darling-Hammond & Youngs, 2002). More recently, results from international tests of academic achievement have shown that American student achievement in mathematics is much lower than student achievement in many European and Asian countries (National Center for Education Statistics [NCES], 2010). Therefore, researchers have become interested in factors, such as teacher beliefs about mathematics and mathematics self-efficacy, that can enhance student achievement in areas like mathematics.

It has long been believed that teacher's beliefs affect their practice. Ernest (1989) began the conversation about a teacher's belief system by describing a mathematics teachers' belief system as being comprised of three parts: ideas about mathematics as a subject, ideas about the nature of mathematics teaching, and ideas about mathematics learning. There is also a notable relationship between a teacher's self-confidence for teaching mathematics and student's self-confidence as mathematics learners (Stipek et al., 2001). It has also been shown that a teacher's beliefs about mathematics as a discipline can be different from their beliefs about mathematics as a school subject. A teacher's relationship with mathematics as subject matter has influence over their beliefs about mathematics as a discipline and whether or not they can extend mathematical thinking beyond the school subject they teach. Likewise, their experiences teaching mathematics influences their beliefs about mathematics schooling and the processes related to performing mathematical tasks (Beswick, 2012). Phelps (2010) also concluded similarly that preservice teachers' motivational profiles are constructed through multiple sources such as: past performance, vicarious experiences, career goals, views of mathematics as a subject and views of mathematics in teaching. Elementary preservice teachers were motivated by their career goal to be attentive to mathematics content because of the need to be able to teach the content. This relationship between their career goals and understanding of mathematics seemed to assist in their formation of mathematics self-efficacy beliefs (Phelps, 2010).

Mathematics Self-efficacy

Mathematics self-efficacy of students has been studied to a great extent by several prominent researchers in the field. Mathematics self-efficacy has been shown to be a motivator for increasing students' mathematics academic achievement and a predictor of mathematics student achievement (Pajares & Miller, 1994; 1995; Schunk, 1991). High mathematics self-efficacy is related to better use of self-regulated learning strategies as well as greater confidence in mathematics ability (Pajares, 1997). It has also been shown to influence career choice (Betz & Hackett, 1983).

Although student mathematics self-efficacy is incredibly important, self-efficacy of mathematics teachers impacts their teaching practices associated with mathematics and their beliefs about mathematics (Stipek et al., 2001). The subject matter knowledge of teachers and their ability to communicate that knowledge are important components of teacher effectiveness (Darling-Hammond & Youngs, 2002). Therefore, investigating teachers' mathematics self-efficacy is an essential step to increasing student mathematic self-efficacy. To answer this need, researchers have been investigating how to measure the self-efficacy of mathematics teachers.

The Mathematics Teaching Efficacy Beliefs Instrument (MTEBI) was created by Enochs, Smith, and Huinker (2000) by revising their earlier published Science Teaching Efficacy Beliefs Instrument (STEBI) (Riggs & Enochs, 1990) to be mathematics-specific. The two constructs in the MTEBI are *personal mathematics teaching efficacy* and *mathematics teaching outcome expectancy*. It should be noted that these are a direct reference to Gibson and Dembo's (1984) TES as well as Bandura's self-efficacy theory. However, Bandura (2006) clearly makes the distinction between *efficacy expectations* and outcome expectations, noting that *outcome expectations* are not a component of self-efficacy, which calls these constructs into question. In addition, although the authors report that the MTEBI is both valid and reliable, they indicate that further work on validation of the instrument is needed. Validity of the MTEBI has only been explored in a limited number of studies (Enochs, Smith, & Huinker, 2000; Kieftenbeld, Natesan, & Eddy, 2010).

Mathematics anxiety. Another interesting facet of the discussion about mathematics self-efficacy is the overwhelming prevalence of an opposing state of mind known as mathematics anxiety. Mathematics anxiety is the state of nervousness and discomfort brought upon by the presentation of mathematical problems (Aiken, 1970). Mathematics anxiety is a highly prevalent problem for students and especially for pre-service teachers, who are still students themselves (Hoffman, 2010). Both pre-service and in-service teachers have reported their own mathematics anxiety as a major concern (Bursal & Paznokas, 2006). Students and teachers with high mathematics anxiety perceive that they are less confident, regardless of ability (Ashcraft & Moore, 2009). One would assume that the presence of mathematics anxiety has a negative impact on mathematics self-efficacy beliefs. Mathematics anxiety in teachers is also related to their pedagogical practice (Steele, 1997), in that it impacts their perceived and actual teaching effectiveness.

Self-efficacy has been shown to moderate the effects of mathematics anxiety because self-efficacy is related to superior performance (Hackett, 1985). Therefore, if a teacher's self-efficacy can be increased, mathematics anxiety could be decreased. Hoffman (2010) found that higher levels of self-efficacy were related to enhanced problem solving efficiency in pre-service teachers. Self-efficacy therefore has a negative relationship with mathematics anxiety.

Teachers' Mathematics Self-efficacy Beliefs

Mathematics self-efficacy of teachers is a complex construct, comprised of several components. Because one of the problems with measuring self-efficacy is the inability to clearly define constructs across the literature (Pajares, 1997), a distinction has to be made here between a *teachers' mathematics self-efficacy* and their *self-efficacy for teaching mathematics*. *Teachers' mathematics self-efficacy* refers to a teachers' own belief in their ability to perform mathematical tasks (Kahle, 2008). Teachers' mathematics self-efficacy has been measured in several ways. Hoffman (2010) used a self-report measure where pre-service teachers rated eight mathematical problems on an 11-point scale with 1 being "no confidence at all in solving" and 11 being "total confidence in solving". Betz and Hackett (1983) created The Mathematics Self-Efficacy Scale (MSES) to measure mathematics self-efficacy by asking the participant to rate their confidence in their ability to solve mathematical problems. The MSES was later revised (MSES-R) by Kranzler and Pajares (1997). In contrast to teachers' mathematics self-efficacy, s*elf-efficacy for teaching mathematics* is a teacher's beliefs regarding their ability to teach others mathematics (Kahle, 2008). This construct is measured by instruments like the MTEBI. Few instruments measuring this construct have been published as a majority of studies use the TSES or MTEBI for this purpose.

Issues with Current Measures of Teacher Self-Efficacy

There are fundamental issues with each of the instruments previously discussed, according to the guidelines for measuring self-efficacy that are touted by Pajares (1997) and Bandura (2006). Furthermore, efforts to measure teacher self-efficacy have become theoretically confused (Haines et al., 1980; Henson, 2001). Historically, most of the instruments that have been created to measure teacher self-efficacy with regard to mathematics fall into the categories of: *general teacher self-efficacy, self-efficacy for teaching mathematics,* or *mathematics self-efficacy*. With regard to general teacher self-efficacy, although the TSES has passed several tests of its validity, its construct validity was measured concurrently and has not been tested using confirmatory factor analysis techniques. It was also only tested during initial development with pre-service teachers, although the audience for the instrument is in-service teachers. The TSES also only measures the underlying construct of teacher self-efficacy in a general manner disregarding specific sub-tasks associated with teaching. To this point, even the authors of the TSES, Tschannen-Moran and Woolfolk Hoy (2001), have postulated that teacher efficacy may vary depending on school and classroom context.

Although capturing context-specific teacher self-efficacy has been attempted (Bandura, 2006; Pajares, 1997), some still argue that measures currently in use for doing so are not completely valid (Swackhamer, 2010). One of the issues with context and content-specific measures of teacher self-efficacy is how to balance specificity and generality. Imbalanced instruments cause measures that are too specific to lose predictive power (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998; Tschannen-Moran & Woolfolk Hoy, 2001). Having predictive power is important since most current measures of teacher self-efficacy are used as predictors of outcomes such as student performance, although prediction using self-efficacy should be undertaken with caution (Pajares, 1997).

Another measurement issue is item wording. The STEBI and the MTEBI, both use wording more theoretically aligned with Rotter's theory than with Bandura's (i.e. "When a low-achieving child progresses in mathematics, it is usually due to extra attention given by the teacher"), although the authors claim to adhere to Bandura's theory of self-efficacy. The TES (Gibson & Dembo, 1984) touts to be theoretically aligned to Bandura's theory of self-efficacy, but in reality is more aligned to Rotter's theory because it used the RAND items as a basis for creating new items (Henson, 2001). Some items on the TSES have wording more aligned to locus of control or attribution theory than self-efficacy (i.e. "How much can you do to calm a student who is disruptive or noisy?"). Other instruments use wording indicating the level of "confidence" that a person has for conducting and completing a task although mere self-judgments of general confidence in performing a task are not accurate measurements of self-efficacy for that task (Bandura, 2006; Pajares, 1997). The wording "can" should be used instead of "will" or "confident" to indicate a judgment of perceived ability instead of intention (Bandura, 1996). Dellinger et al. (2008) determined that items with the wording "My belief in my ability to…is" had different results when comparing responses to items worded "I can" or "I am able to." Therefore, the wording of items intended to measure self-efficacy should be carefully considered to ensure proper measurement of the construct.

Another caveat of measuring self-efficacy beliefs is the fact that best practice calls for measurement of self-efficacy at the most optimal levels of specificity relevant to the criteria task being assessed (Bandura, 2006; Pajares, 1997). Pajares (1997) indicated that in order to properly measure self-efficacy, the specific task in question has to be identified first. Then, the levels of that specific task have to be identified by creating items that relate to the difficulty of the sub-tasks for each level. If the task is already learned, task-specific items are relevant. If the task is novel, language about perceived self-efficacy for that task is more appropriate. This difference between learned tasks and learning future tasks has created a distinction between “self-efficacy for performance” and “self-efficacy for learning” in the literature (Pajares, 1997).

Along with item wording issues and context issues, the overall use of self-efficacy measures has been called into question (Pajares, 1997). Because rigorous construct validity testing of new teacher self-efficacy instruments is often not in the interest of researchers, most researchers use already existing instruments when conducting teacher self-efficacy research (Henson, 2001; Pajares, 1997). Researchers often choose a self-efficacy instrument without understanding the constructs it intends to measure. Most self-efficacy instruments use a total score as a predictor of student achievement (Pajares, 1997), a practice that Bandura (1997) strongly cautioned against because of his belief in context-specific self-efficacy. Other researchers have found that sub-scores for the constructs are more appropriate with some groups (pre-service teachers) instead of using a total score (Tschannen-Moran & Woolfolk Hoy, 2001). The use of the instrument should match the theoretical framework of the research and the theoretical underpinnings of the instrument.

Currently most instruments are created to exclusively measure either *general self-efficacy*, *mathematics self-efficacy* or *self-efficacy for teaching mathematics*, ignoring the potential correlations between theses constructs. This may not be the best practice for determining a teacher’s true beliefs about mathematics teaching and learning as Ernest (1989) asserts that a mathematics teacher’s belief district has three parts: ideas about mathematics as a subject for study, ideas about the nature of mathematics teaching, and ideas about mathematics learning. Also, the lack of cohesion within the field of a definite factor structure that defines self-efficacy leads to the diversity of self-efficacy instruments in the literature (Pajares, 1997). Therefore, it is logical that an instrument intended to measure the self-efficacy of mathematics teachers needs to be more comprehensive.

Finally, the construct validity of existing self-efficacy instruments has been called into question. The TES, by most factor-analytic standards, has poor factorial validity (Henson, 2001). The variance accounted for by the two factors cited in Gibson and Dembo (1984) is only 28.8% total (Henson, 2001). Denzine, Cooney, & McKenzie (2005) also conducted a confirmatory factor analysis on the TES, which showed that the two dimensions suggested to be present in the instrument did not fit the data. A recent re-examinination of the MTEBI using Item Response Theory (IRT) provided evidence that the validity of the MTEBI was not as high as previously thought (Kieftenbeld, Natesan, & Eddy, 2010). Even the TSES, the most widely accepted instrument, has only been re-examinined a few times with regard to validity since its original publication in 2001 (Heneman, Kimball, & Milanowski, 2007; Klassen et al., 2009) although it is used widely in studies that examine elements of teacher self-efficacy (Wolters, & Daugherty, 2007; Ross & Bruce, 2007; TSchannen Moran & Woolfolk Hoy, 2005, and others). The TSES constructs do seem to hold true over time and in different contexts (Heneman, Kimball, & Milanowski, 2007; Klassen et al., 2009) but not always when modified to be mathematics-specific (Swackhamer, 2010).

Developing and Validating an Instrument

Since this study concerns the validation of an instrument designed to measure the self-efficacy of elementary mathematics teachers, some discussion on related literature is important. There are several facets of the knowledge base concerning instrument construction and validation that must first be understood. First, to construct an instrument it is vital to understand the mechanism for the delivery of items and the items themselves. These skills are a delicate balance between psychometric understanding of the cognitive processes that one is attempting to capture through item creation and the construction of an instrument in a way that captures measurable correlates to those internal processes. The understanding of a conceptual and theoretical framework is crucial to creating a psychologically based instrument. Secondary to this is a deep understanding of the procedures that can be used to show evidence of reliability and validity. These evidences are based firmly in statistical procedures but cannot be conducted without first having a clear understanding of instrument construction and delivery.

Survey Research

Survey research is a necessary method for gathering data in educational research because it has the ability to allow researchers to study complex topics (Ebel, 1980). Surveys utilize self-report, a psychologically sound method of collecting first-hand data from an individual, especially when observation is not enough (Baldwin, 2009; Harrell, 1985). Although creating a survey may seem simple, a technically-adequate survey instrument must pass several rigorous tests to be considered useful for the purpose in which it was intended. Survey instruments must show evidence of reliability and validity (Johnson & Christensen, 2004).

Validation: Reliability

An important first step in validation of an instrument is to determine its stability, or reliability. Reliability is the consistency of measurement when the testing procedure is repeated on a population (AERA, 1999; Colosi, 1997; Johnson & Christensen, 2004). The statistical coefficient used to indicate reliability is Cronbach's Alpha. A high Cronbach's alpha ensures that the interpretation of the instrument data can be trusted to remain consistent across time and administrations. There are numerous types of reliability, including internal consistency. Internal consistency implies that the questions intended to measure the same concept do so on a consistent basis and in a consistent way (Colosi, 1997).

Validation: Validity

The simplest definition of validity would be to describe it as "accuracy." Validity takes different forms because there are different ways that scores can be accurate (Huck, 2008). It is the interpretation of the test score that is validated, not the test (Messick, 1995). There is not on overarching statistical procedure or coefficient for determining validity. Instead, evidence of validity is shown by collecting evidence of the accuracy of measurement of the instrument, and thus the accuracy of the score (Messick, 1995). Messick (1995) explores six aspects of validity in his seminal work: content, substantive, structural, generalizability, external, and consequential. The *Standards for Educational and Psychological Testing* (AERA, 1999) also state that evidence of validity should be based on elements such as: test content, response processes, internal structure, relations to other variables, and the consequences of testing. The sources of validity evidence needed may differ depending on the use of the test or scale and the interpretations of the test scores (AERA, 1999).

Content validity is of utmost importance. It is established deductively by defining what one intends to measure and then creating items that sample that content and is determined by the judgment of those considered to be experts in the content area to be tested (Cronbach & Meehl, 1955). There is no quantitative measure for content validity, only the opinion of experts as to how the instrument is constructed and how the items represent the content area (Gay et al., 2009; Huck, 2008).

Construct validity is paramount to providing evidence that an instrument actually measures what it is intended to measure. A construct is a set of attributes that a person is assumed to possess or possesses in varying degrees (Cronbach & Meehl, 1955). Often constructs have psychological underpinnings as they might be developed to describe an internal process (Messick, 1955). Determining construct validity requires gathering a lot of evidence (Messick, 1955). It is important that a good definition and meaning of the construct to be tested is understood and operationally defined (Johnson & Christensen, 2004). A researcher can use several methods to provide evidence of construct validity including: providing correlational evidence between certain measured variables that have conceptual relationships to the variables being measured in the new instrument, showing the difference in scores between groups of respondents, or conducting factor analysis on scores from the new instrument (Huck, 2008). As new information about the validity of a test is discovered, changes to the test, its underlying constructs, and conceptual framework might be needed (AERA, 1999).

Using Statistical Procedures to Collect Evidence for Validity

When examining relationships within aspects of construct validity, Pearson's product-moment correlation coefficient (*r*) is one correlation coefficient that can be used (Huck, 2008; Schumaker & Lomax, 2010). The correlation coefficient is reported as a decimal number between -1.00 and +1.00. The indication of a high or low correlation varies depending on the field of study, research questions, and context. Caution should be used when evaluating correlation coefficients so strictly as the purpose of the study influences the interpretation of the coefficient (Huck, 2008). It is common to use the square of the correlation coefficient (r^2) to show the size of the relationship instead (Ozer, 1985). This index is the coefficient of determination and is interpreted as the percentage of variance of one variable explained by the other variable (Ozer, 1985).

Factor analysis techniques provide insight into the theoretical constructs represented by questions on a measure. It is assumed that items, or observed variables, outnumber the theoretical constructs represented. The sets of observed variables that are clustered into a construct are determined empirically by the shared variance-covariance among observed variables. Data are collected for observed variables and then factor analysis is conducted to either explore factor structure or confirm factor structure (Schumaker & Lomax, 2010).

An EFA is usually conducted in the early stages of research when it can serve as a tool for consolidating variables and examining the underlying factor structure of an instrument (Tabachnick & Fidell, 2007). In Exploratory Factor Analysis (EFA), the underlying factors in a set of data are explored in order that a factor structure, or model, can be uncovered. An EFA isolates the structure of potential factors mathematically, although interpretation of the factor structure is dependent on the theoretical framework of the research (Thompson & Daniel, 1996).

Once data are obtained, several steps must be taken to conduct an EFA. A correlation matrix is used to determine the relationship between observed variables (DeCoster, 1998). The correlation is computed using variances and covariance input into the following formula (Stevens, 1996):

$$r = \frac{1}{n-1}\sum\left(\frac{\chi-\overline{\chi}}{S_{\chi}}\right)\left(\frac{y-\overline{y}}{S_{y}}\right)$$

(*r* = linear correlation coefficient, n = sample size, $\sum$ = sum, *x* = mean, $\overline{\chi}$ = sample mean, S_{χ}= standard deviation of *x*, *y* = mean, $\overline{y}$= sample mean, S_{y}= standard deviation of *y)*

The researcher can specify the number of factors present and allow the data to drive the factor structure of the observed variables or the number of factors can be set *a priori* (Schumaker & Lomax, 2010). Principal Component Analysis is one technique for extracting factors. It works by extracting the dominant patterns in the matrix according to a set of loading points (Wold, Esdensen, & Geladi, 1987). To extract factors, the factor pattern coefficient matrix is multiplied by the factor correlation matrix to yield the factor structure matrix (Thompson & Daniel, 1996). An eigenvalue is a statistic used, in this case, to inform the researcher about variance accounted for before the factors are rotated (Thompson & Daniel, 1996). An eigenvalue is the root of a *p* x *p* matrix A. It is the solution to the following equation:

$$|A - \lambda I] = 0$$

(A is a square matrix, λ (lambda) = scalar)

A will have *p* roots, some of which may be 0 (Stevens, 1996). Once eigenvalues are obtained, factors are then rotated to find a final solution. Orthogonal rotation is used to produce uncorrelated factors, oblique rotations produce correlated factors (DeCoster, 1998). If the factors are correlated, direct oblimin can be used (DeCoster, 1998). Structure coefficients are also referred to as "loadings" in some literature (Thompson & Daniel, 1996). Comrey and Lee (1992) suggest that loadings in excess of .71 are considered excellent, .63 very good, .55 good, .45 fair, and .32 poor. Loadings less than .32 (less than 10% overlapping variance) are usually not interpreted.

Another type of factor analysis is Confirmatory Factor Analysis (CFA). In a CFA, the intention is to test data to confirm an already existing theoretical model. CFA is also useful for testing rival models. In CFA, the researcher specifies the factor structure and relationships in the CFA technique. In CFA, additional data can be continuously tested against the hypothesized model and if the data fits the model this further validates the model (Schumaker & Lomax, 2010). A CFA is usually conducted later in advanced stages of the research process. It is a technique that is more complicated and is usually performed through structural equation modeling (Tabachnick & Fidell, 2007).

Shumaker and Lomax (2010) suggest a five-step model for conducting CFA: model specification, model identification, model estimation, model testing, and model modification. When a model is specified, and subsequently tested, the software (e.g., LISREL, AMOS, EQS) can use raw data, a correlation matrix, or a variance-covariance matrix to determine how well the data fit the model. If a correlation matrix is input into the computer program, it is most likely converted into a variance-covariance matrix using the standard deviations of the variables (Schumaker & Lomax, 2010).

When determining model fit, there isn't one lone fit statistic that is used. Therefore, it is suggested to report several (Thompson & Daniel, 1996). Schumaker and Lomax (2010) suggest that the following be reported: χ^2, degrees of freedom (*df*), p value (*p*), root-mean-square error of approximation (RMSEA), and goodness-of-fit index (GFI). The RMSEA should be below .07 and the GFI should be above .95 (Schumaker & Lomax, 2010). The χ^2 value should be relatively close to the *df* in order for the model fit to be good. Important to note is that the χ^2 statistic is essentially "badness of fit" and if it is statistically significant, modifications to the model should be made (Schumaker & Lomax, 2010).

Hu and Bentler (1999) also suggest certain cut-off values for determining model fit. Various fit indices can be used that are MLE-based: Tucker-Lewis Index (TLI), Bollen's Fit Index (BL89), Relative Noncentrality Index (RNI), Comparative Fit Index (CFI), or Gamma Hat. The cut-off value for these suggested by Hu and Bentler (1999) is .95. In order to reduce Type II error, Hu and Bentler also suggest a cutoff for the RMSEA of close to .06. When determining goodness of fit it is important to examine several fit indices and not look at one lone statistic. Hu and Bentler (1999) also suggest using a cutoff value of .95 for TLI (or BL89, RNI, CFI, Gamma Hat) in combination with a cutoff value of close to .09 for standardized root mean squared residual (SRMR) to evaluate model fit as it resulted in the least sum of Type I and II error rates.

Combinational rules with RMSEA of greater than .05 (or .06) and SRMR greater than .06 also show acceptable model fit with reasonable Type II error rates. The least sum of Type I and II error rates can be found using a combination of RMSEA greater than .06 and SRMR greater than .09 (Hu & Bentler, 1999). However, caution should be used when the sample size is less than 250 because the combinational rules have a tendency to over-reject true population models (Hu & Bentler, 1999).

Although these general rules set forth by Hu and Bentler (1999) and Schumaker and Lomax (2010) do provide some simplicity, the two-index strategy has questionable validity (Fan & Sivo, 2005). Fan and Sivo (2005) found that the SRMR fit index is not necessarily the most sensitive to misspecified factor covariances and the group of fit indices used by Hu and Bentler are not more sensitive to misspecified factor loadings (Fan & Sivo, 2005). Marsh, Hau, and Wen (2004) also caution that Hu and Bentler's suggestions for using certain combinations of cutoff values for certain fit indices have been adopted by the profession in an attempt to find some sort of "golden rule" for goodness of fit. It is problematic to adopt the suggestions by Hu and Bentler as an absolute rule without taking into account both their intentions for publishing guidelines and the individual researcher's own theoretical foundations and knowledge of their data. Adopting Hu and Bentler's overly strict guidelines could cause the rejection of some otherwise seemingly acceptable models. Therefore, greater consideration should be given to the overall model fit without strict attention to certain cutoff values for fit indices (Marsh et al., 2004).

The final step in Shumaker and Lomax's (2010) five step model for CFA is model modification. Linear Structural Relations (LISREL), statistical software used to conduct a

CFA, often suggests modifications to the model to improve model fit by either adding paths from observed to latent variables or by adding error covariance between observed variables. MacCallum, Roznowski, and Necowitz (1992) caution against using the software's suggestions without regard for theoretical underpinnings of the model. Modified models may not cross-validate consistency well unless large sample sizes are used. Additionally, paths should not be added to the model in an attempt to improve model fit if there is no logical theoretical basis present (MacCallum et al., 1992).

Once construct validity is established, external aspects of validity can be explored (Messick, 1995). The external aspect of validity can be determined by examining discriminant validity as well as concurrent and criterion-related validity. The purpose of discriminant validity is to show that there is no statistically significant relationship between the construct measured on the new instrument and scores on an established instrument that measure theoretically different constructs (Johnson & Christensen, 2004). The negative relationship is theoretically sound and important evidence of validation (Huck, 2008). The general purpose of discriminant validity is to simplify the theoretical constructs present in the literature. Using discriminant validity as a way to indicate the difference between two constructs allows for each newly-introduced construct to be examined separately (Johnson & Christensen, 2004). Concurrent validity is established by correlating scores on the instrument to scores on an existing instrument. The degree to which new instruments can be deemed accurate measures of a variable is by comparing scores from the new instrument to scores on a relevant criterion variable (Huck, 2008). If the resulting correlation is high, the instrument has acceptable concurrent validity.

Concurrent validity can also be examined if a new test is administered within a short period of time of an existing test to be sure that scores are correlated (Huck, 2008).

Rationale and Purpose for this Study

In an effort to provide more context-specific measurement of a teacher's self-efficacy beliefs, the focus of this study is the development of one such instrument; the Self-efficacy for Teaching Mathematics Instrument (SETMI). The SETMI is a survey instrument that aligns to the theoretical underpinnings of Bandura's Social Cognitive Theory and also the idea that a teacher's belief system surrounding mathematics is complex (Ernest, 1989). Although several instruments have been used in an attempt to measure self-efficacy for teaching mathematics, most either fail to provide substantial evidence of construct validity (Swackhamer, 2010) or use over-generalized statements out of context of the criteria task (Pajares, 1997). Other researchers have used *locus of control* instead of *self-efficacy* constructs to measure self-efficacy, and this has not proven fruitful either (Henson, 2001). Another issue with current measures of teacher self-efficacy, including the TSES, is the lack of specificity to a teacher's grade level (elementary, middle, or high school). Because education is different at each of these grade levels, expectations of a classroom teacher and sub-tasks associated with teaching content are different. In an effort to measure context-specific self-efficacy, elementary education is the setting of this study.

Elementary teachers in the United States are often expected to be able to expertly teach all subjects to their students, but this may not be how they view their abilities. Some elementary teachers have communicated their apprehension to researchers and teacher educators about mathematics and their own self-doubts regarding teaching mathematics (McGee, Wang, & Polly, in press; Piel & Green, 1993; Polly, McGee, Wang, Lambert, Pugalee, & Johnson, 2010). Because self-efficacy is context specific, teachers often have varied beliefs toward different aspects of teaching mathematics (Ernest, 1989). Some teachers feel highly efficacious in their knowledge of content but less efficacious in their ability to teach those concepts to students. Others have content-specific self-efficacy beliefs and may feel extremely efficacious about teaching some aspects of the content (i.e. patterns) but less efficacious about teaching other content (i.e. fractions) (McGee, Wang, & Polly, in press).

This study considers the elementary mathematics content and context with regard to teachers' self-efficacy. The purpose of this study is to show evidence of reliability and validity for a survey instrument that measures the mathematics self-efficacy of elementary teachers. The following research questions are of interest:

1. What is the reliability of the newly-developed instrument?
2. What is the structure of the newly-developed instrument?
3. What evidence exists regarding validity of the instrument?

Chapter Summary

The concept of self-efficacy began with discussions in the literature regarding Rotter's (1954) Social Learning Theory and continuing with Bandura's Social Cognitive Theory (1977). Before the publication of Bandura's theory, two items on RAND study evaluations alluded to teacher self-efficacy by asking teachers to rate their perceived ability to serve students. These two items used Rotter's theory as a foundation, citing locus of control as the theoretical basis behind early thoughts about teacher self-efficacy. Later studies adopted Bandura's theory as a basis for their work as it included ideas about instincts and internal motivation that were not included in Rotter's theory. In fact, Rotter himself often disagreed with the interpretation of locus of control in academic work and often wrote about the subject. Bandura's work has thus led the charge for work on teacher self-efficacy for the last 30 years.

In an effort to better understand teacher self-efficacy, researchers have created survey instruments and relied on self-reporting of teachers. The range of survey instruments varies widely and includes both likert-type items and scenarios as an attempt to capture the "elusive construct" (Tschannen Moran & Woolfolk Hoy, 2001) of self-efficacy. Some of the first instruments published were regarded as being highly valid and quickly adopted into practice, while others faded out of the literature over time. The creation of so many instruments in such a short period of time led to confusion over the definition of teacher self-efficacy, its associated constructs, factor structures, and appropriate measurement. Bandura himself created an instrument in an attempt to quell the debate over self-efficacy measurement, but unfortunately the instrument was never published.

The cause of teacher self-efficacy research has been led by Tschannen Moran and Woolfolk Hoy in recent decades. These researchers have successfully created and validated an instrument to measure general teacher self-efficacy using Bandura's work as a theoretical and practical basis (2001). The TSES is now considered the most accurate instrument for the measurement of general teacher self-efficacy and is widely used in the literature.

Other content-specific instruments exist, such as the STEBI and MTEBI (Enochs, et al., 2000; Riggs & Enochs, 1990). These two instruments specifically are used quite often in the research, but closer exploration of the items reveals a lack of specificity of math and science content and a potential conflict with Bandura's theory. With regard to content-specific self-efficacy, both Bandura (1994) and Pajares (1997) explain that measuring self-efficacy within specific contexts is the best practice. Both researchers also note the importance of finding balance between specificity and generalizability with regard to instrumentation.

Generally speaking, content-specific instruments have been difficult to validate and efforts to do so ended in the early 2000's. Most research to date cites either the STEBI or MTEBI as measures of self-efficacy in science and math and researchers seem to have lacked the desire to create and validate newer measures. Another issue with the context-specific instruments mentioned above is the lack of specificity of sub-tasks associated with teaching content and specific levels of education (elementary, middle, and high school). Schooling in the United States has become extremely focused on student achievement at each level of schooling, thus creating distinct cultures within classrooms at each level. In effect, the mathematics self-efficacy of an elementary teacher might differ greatly for the mathematics self-efficacy of a high school teacher.

In an effort to measure content and context-specific self-efficacy, a newly-created instrument will be used as the subject of this study. This newly created instrument is both mathematics-specific and specific to the context of teaching mathematics in an elementary setting. The purpose of this study is to provide evidence for the reliability and validity of this instrument.

CHAPTER 3: METHODOLOGY

The purpose of this study was to show evidence of reliability and validity for a survey instrument to measure the mathematics self-efficacy of elementary teachers. This chapter highlights the methodological procedures relevant to establishing reliability and validity evidence for the new instrument. The proposed theoretical framework will be discussed as well as the development of the survey itself. Participant demographics along with phases of data collection will also be discussed. In an effort to provide comprehensive evidence for the reliability and validity of the instrument, the following research questions were of interest:

1. What is the reliability of the newly-developed instrument?
2. What is the structure of the newly-developed survey instrument?
3. What evidence exists regarding validity of the instrument?

Theoretical Framework

This study considers the nature of teaching elementary mathematics with regard to a teacher's mathematics self-efficacy. As mentioned in Chapter 2, there exist only a few valid and reliable instruments for measuring teachers' self-efficacy. The most widely used and accepted instrument for this purpose is the Teacher's Sense of Efficacy Scale (TSES) (Tschannen-Moran & Woolfolk Hoy, 2001). In the pursuit of measuring content-specific self-efficacy, some researchers have created instruments to measure teacher self-efficacy in mathematics and science (Enochs, Smith, & Huinker, 2000; Riggs & Enochs, 1990), but those instruments only vaguely align with Bandura's Social Cognitive Theory (Bandura, 1977) and fail to include specificity of sub tasks associated with teaching mathematics (Pajares, 1997). Other self-efficacy instruments that have been created since teacher self-efficacy began to be studied either failed to be accepted by the profession or failed to be valid over time (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998; Swackhamer, 2010). Therefore, a new instrument has been created to measure elementary teachers' mathematics self-efficacy and the evidence for validity of the instrument is presented in this study.

Context

Data collection for this study occurred during the second and third year of a three-year federally-funded Mathematics Science Partnership (MSP) grant. This project was funded by a $2.4 million dollar grant from the United States Department of Education, through the state Department of Public Instruction. The MSP grant project took place between August 2009 and June 2012. The MSP grant project was a collaborative partnership between two school districts and the university where the researcher is a student. District One is a large urban district with 88 elementary schools and District Two is a neighboring suburban district with five elementary schools.

Each year of the MSP grant project, teachers applied to be selected as participants in the MSP grant through procedures defined by their school districts. District One teachers were selected based on the number of applicants at their school, with attention paid to selecting large numbers of teachers at each school. District Two teachers were selected in an effort to have all teachers eventually participate in the grant by the end of the third year. On average District One selected 200 teachers per year and District Two selected 30.

Selected teachers participated in 84 hours of professional development, which included a 60-hour summer institute followed by 4, 6-hour follow-up sessions during the school year. Professional development in this 60-hour institute was conducted primarily by a mathematics professor and an elementary education mathematics professor at the partnering university. Follow-up sessions in years two and three were led by teacher-leaders who were participants in Cohort I of the MSP grant. Teachers from the two school districts participated in the professional development separately and on different days, but the overall content and focus of the professional development remained consistent.

The purpose of the MSP grant project was to develop standards-based elementary mathematics educators by providing intensive professional development and support centered on the *Investigations in Number, Data, and Space* curriculum (TERC, 2009). Subsequent goals of the MSP grant program were to increase teachers' mathematics content knowledge, improve student learning outcomes, create a cohort of teacher-leaders who will lead school based professional development on standards based math curriculum, and to create a successful model of professional development. Within the framework of the MSP project goals, attention was given to the influence of professional development on teacher's belief systems and the importance of understanding and shifting teacher's beliefs in order to aid them in implementing new teaching practices. Measuring teachers' self-efficacy for teaching mathematics provided a unique view of a teacher's complex belief system. Data used in this study was a portion collected as part of the program evaluation. Cohort II (2010-2011) and III (2011-2012) data were used in this study.

Participants

As noted earlier, three cohorts of teachers were involved in the MSP grant during the three years in which it was implemented. Those three cohorts are as follows: Cohort I (2009-2010), Cohort II (2010-2011), and Cohort III (2011-2012). Cohort II of the MSP grant (2010-2011) provided the data for Phases I and II of this study. There were 183 teachers in Cohort II, 153 taught within a large urban school district (District One) and 30 taught in a much smaller district located in a nearby city (District Two). One hundred seventy one of the participants were female (93%) and 12 were male (7%). One hundred and three (56%) of the participants were Caucasian, 34 were African American (19%), two were Hispanic, one Asian, and two self-identified as Other. Twenty-nine (16%) of the participants taught Kindergarten, 26 (16%) taught first grade, 29 (16%) taught second grade, 34 (19%) taught third grade, 26 (14%) taught fourth grade, and 36 (20%) taught fifth grade. Three were Exceptional Children (EC) teachers who assisted other teachers but did not teach their own mathematics classes. All of these Cohort II participants completed the Self-efficacy for Teaching Mathematics Instrument (SETMI) in August 2010 during the summer institute and again in February 2011.

Cohort III of the MSP grant (2011-2012) provided data for Phases III and IV of this study. There were 228 teachers in Cohort III. One hundred ninety were from District One and 36 were from District Two. Thirty-seven (17%) of the participants taught Kindergarten, 42 (19%) taught first grade, 37 (17%) taught second grade, 39 (17%) taught third grade, 39 (17%) taught fourth grade, and 23 (10%) taught fifth grade. Six teachers were Exceptional Children (EC) teachers who assisted other teachers but do not teach their own mathematics classes. Ethnicity and gender information for Cohort III had not been collected at the time this study was conducted. All Cohort III participants completed the Self-efficacy for Teaching Mathematics Instrument (SETMI) in August 2011 during the MSP summer institute. Table 2 highlights participant demographics for both Districts.

Table 2.

Participant Demographics

	Cohort II	Cohort III
District One:	*n* = 155	*n* = 243
Kindergarten	20 (12.9%)	33
1st Grade	23 (14.8%)	48
2nd Grade	25 (16.1%)	42
3rd Grade	29 (18.7%)	42
4th Grade	26 (16.8%)	47
5th Grade	32 (20.6%)	31
Males	13 (8.4%)	n/a
Females	142 (91.6%)	n/a
Caucasian	78 (66.1%)	n/a
African American	35 (22.6%)	n/a
Hispanic	2 (1.3%)	n/a
Asian	1 (0.6%)	n/a
Other	2 (1.7%)	n/a
Unspecified	37 (23.9%)	n/a
Math Certification	64 (41.3%)	n/a
District Two:	*n* = 30	*n* = 32
Kindergarten	9 (30%)	1
1st Grade	4 (13.3%)	3
2nd Grade	4 (13.3%)	7
3rd Grade	2 (6.7%)	7
4th Grade	4 (13.3%)	4
5th Grade	5 (13.3%)	4
Exceptional Children's (EC) teachers	3 (10 %)	6
Males	0	n/a
Females	30 (100%)	n/a
Caucasian	29 (96.7%)	n/a
African American	1 (3.3%)	n/a
Math Certification	14(46.7%)	n/a

Research Design

This study took place in four phases. Phases I and II used data collected from Cohort II of the MSP grant. In Phase I of this study, the participants answered questions on the newly-developed instrument as a pre-test in August 2010. One hundred eighty three instruments were distributed. Data from this phase of data collection was used to determine reliability of the new instrument by determining the correlation coefficient of the items within each construct. This data was also used to explore the structure of the constructs in the instrument using Exploratory Factor Analysis (EFA) and to explore concurrent validity by correlating the new instrument to the TSES.

In Phase II of this study, Cohort II participants responded to items on the same, unaltered instrument as a post-test measure in February 2011. One hundred and eighty three instruments were distributed. These data were used to provide evidence of stability by correlating each participant's pre-test score to their post-test score on the new instrument.

Phases III and IV of this study took place during year three of the MSP grant project (2011-2012). In Phase III, the researcher met with an elementary mathematics expert from the university in June 2011 to revise mathematics content items on the SETMI Version One in order that they aligned better to the state standard course of study. Items were also revised to clarify the content being assessed and to shorten the survey length if possible. The basis for this change was also influenced by output from an Exploratory Factor Analysis (EFA).

Afterwards, a focus group of five elementary teachers from Cohort I of the MSP grant met in July 2011 to discuss the new items on the instrument. A logical initial step in developing an instrument for a specific population is to test the instrument with members of that population through pilot testing. A pilot test can consist of as few as three or four people completing the questions and providing feedback. These testers, or focus group members, should provide feedback about commission and omission as well as how items are interpreted. Having the focus group examine the completeness of the survey is one way to examine content validity (Gay et al., 2009). The focus group provided feedback to the researcher regarding item wording. Only one item was suggested to be revised.

Cohort III participants responded to items on the instrument as a pre-test measure for MSP grant purposes in August 2011 (Phase IV). Two hundred twenty-two instruments were distributed. This set of data was used to conduct a Confirmatory Factor Analysis (CFA) on the final version of the Self-efficacy for Teaching Mathematics Instrument. This analysis will complete Phase IV of this study.

Instrument Development

Within the literature there exist both general and context specific instruments exist that are intended to measure teacher self-efficacy. The most widely accepted measure of general teacher self-efficacy, the TSES (Tschannen-Moran & Woolfolk Hoy, 2001), as mentioned in Chapter II, is deemed valid and reliable based on empirical research from several studies. The issue with the TSES for some purposes is that it only measures general self-efficacy and lacks the specificity that is preferred when accurately measuring self-efficacy for a specific task or within a specific context (Pajares, 1997). When searching for mathematics specific instruments to measure teacher self-efficacy the most commonly used measure is the MTEBI (Enochs, Smith, & Huinker, 2000). The MTEBI was deemed valid and reliable by its authors, but fails to align properly with Bandura's theory of self-efficacy, the self-efficacy theory most widely prescribed to in the literature. The MTEBI and other contextual self-efficacy instruments also fail to be specific to a teacher's grade level. Therefore, the purpose of this study was to examine evidence of validity for a newly created instrument that measures the mathematics self-efficacy of elementary teachers entitled the, “Self-efficacy for Mathematics Teaching Instrument (SETMI).” The SETMI was first created in July 2010

The SETMI was developed using two instruments as a framework for the creation of items. The TSES (Tschannen-Moran & Woolfolk Hoy, 2001) was used to guide work on the SETMI because it is the most widely accepted measure of general teacher self-efficacy. The short form of the TSES contains 12 questions that address three constructs: *efficacy in student engagement*, *efficacy in instructional strategies*, and *efficacy in classroom management*. Another instrument developed by Aerni (2008) entitled “Teaching Mathematics in Inclusive Settings was also used as a guide for the SETMI. The “Teaching Mathematics in Inclusive Settings” instrument uses the TSES short form, modified to be specific to teaching mathematics in inclusive settings. It also contains mathematics content items.

Because classroom management is not one of the purposes of the MSP grant, those items from the TSES were not used in any form on the SETMI (TSES items 1, 3, 6, 8, and 11). In addition, the creators of the TSES (Tschannen-Moran and Woolfolk Hoy, 2001) acknowledge that *classroom management* was not a construct they originally considered to be part of self-efficacy for teaching but was a third construct that was added through subsequent validation studies because some of their focus groups suggested that those items be added. The questions on the TSES associated with this construct are also very specific to managing only disruptive student behavior.

Taking the context of the study and the theoretical framework for self-efficacy into consideration, Part One of the SETMI uses items 2, 4, 5, 7, 9, 10, and 12 from the TSES short form. These items were modified to be mathematics specific. One additional item for SETMI Part One comes from Aerni's instrument (How much time do you feel you have to teach mathematics?) and another was added at the request of mathematics experts who are part of the MSP grant management team (To what extent do you effectively teach mathematics?).

Part Two of the SETMI, *efficacy for understanding elementary mathematics content* (*EUMC*), asks teachers to rate their confidence level for their understanding of mathematics content specific to elementary school (i.e. integers, rational, irrational numbers and size, quantity and capacity). Part Three of the SETMI, *efficacy for teaching mathematics content* (*ETMC*), asks teachers to rate their confidence in their ability to teach students the same mathematics concepts. Parts Two and Three of the SETMI were created using Aerni's (2008) content specific items as a guide, the state Standard Course of Study for elementary school, and the *Investigations* curriculum. Content experts were consulted to be sure that items communicated clearly. Dr. Drew Polly, mathematics teacher educator and Assistant Professor, two elementary mathematics education directors for the large urban school district and a K-4 elementary mathematics facilitator for the smaller suburban school district were consulted as content experts. Any needed revisions to the SETMI were reviewed by a focus group of elementary mathematics teachers. These elementary teachers are members of Cohort I of the MSP grant and were professional development facilitators for Cohort II and Cohort III of the MSP grant project. The review of items by content experts during each phase of this study provided evidence for content validity.

Prior to an Exploratory Factor Analyses, the SETMI was expected to measure three constructs: *mathematics teaching self-efficacy (MTSE) (items 1-9), efficacy for* *understanding elementary mathematics content* (*EUMC*) (items 10 through 24), and *efficacy for teaching mathematics content* (*ETMC*) (items 25 through 39).

Procedures

SPSS version 15 was used for descriptive and inferential analysis of the data. Reliability of the SETMI was determined by calculating Cronbach's Alpha for each of the four original constructs represented. Data for this procedure were used from the August 2010 administration of the SETMI to MSP grant participants. One hundred eighty three instruments were distributed and 183 participants completed the SETMI.

Factor structure of the SETMI was then examined using an Exploratory Factor Analysis (EFA). For participants that chose two responses to an item on the SETMI, the midpoint of the two responses was used as their response (e.g.,. 2 and 3 were selected, 2.5 was computed). Principal Component Analysis (PCA) was utilized in the EFA and VARIMAX rotation was used (Kaiser, 1958). The anticipated number of factors was not set prior to the analysis. Three separate EFA's were conducted because of the structure of the SETMI in three parts. The TSES items were also on a nine-point scale while the items in Parts Two and Three were on a 5-point scale. The nine-point scale comes from the TSES and is from Bandura's literature on self-efficacy scale design (2006).

Concurrent validity was examined by correlating Parts Two (EUMC) and Three (ETMC) of the SETMI to Part One (MTSE). Part One of the SETMI, by containing items from the TSES, has strong, positive, and statistically significant correlations between the two constructs present: *efficacy for student engagement in mathematics* (items 1, 2, 4, and 7) and *efficacy for mathematics instructional strategies* (items 3, 5, 6, 8, and 9) (r = .75). These two constructs combined are second-order factors comprising the construct *mathematics teaching self-efficacy* (*MTSE*). Part One of the SETMI is used in this manner as an already existing, valid measure of teacher self-efficacy because items come directly from the TSES, a valid and reliable measure of teacher self-efficacy. Pearson's *r* was the statistical measure used to ensure concurrent validity is present. See Tables 3, 4, and 5 for item correlations.

Phase II of the study included an evaluation of test-retest reliability. Test-retest reliability was examined by correlating participant responses from two administrations (August 2010 and February 2011). Pearson's *r* was used as the correlation coefficient

Phase III of the study included an examination of possible revisions to the instrument. After the EFA was conducted, the factor structure was analyzed and some revisions were made to the SETMI in an effort to limit the number of constructs in the model. Specific revisions are detailed in Chapter 4 and Tables 8 and 9. An effort was made to keep the instrument at reasonable length while also ensuring that adequate items per construct were maintained. The item scaling on Part One as nine-points was modified to be a five-point scale in Version Two so that the entire instrument was on the same scale.

In Phase IV of this study, a Confirmatory Factor Analysis (CFA) was conducted using data from Cohort III. Two hundred twenty-two instruments were distributed and all were completed. The specified model contained three constructs (MTSE, EUMC, and ETMC) and their relevant observed variables as mentioned above, For other statistical analyses, The error variance of the three latent variables (MTSE, EUMC, and ETMC) was set to be free. The three latent variables were also set as second-order factors under the main factor *self-efficacy for teaching mathematics* (SETMI). For the CFA, LISREL was used. Model modifications that were suggested were only considered if they fell within the specified theoretical framework. Chapter 4 contains all results and statistical output from the procedures outlined in this chapter.

Minimizing Threats to Internal Validity

In order to maintain the integrity of this study, threats to internal validity were minimized. This research spanned two years and included two groups of participants. One threat to the internal validity of this study is the impact of history, or the events that may impact the self-efficacy of the teacher outside of the scope of this study. Participants might have been involved in multiple mathematics professional development projects or might have been required to take additional professional development, both of which could have impacted their self-efficacy for teaching mathematics. In order to minimize this threat to internal validity, the timing for administration of the SETMI for both the EFA analysis and the CFA analysis occurred during the summer and before teachers had participated in any professional development program relevant to the MSP program. Additionally, participants only completed the SETMI two times during their tenure as participants in the MSP grant. Data collected at each time was used for different data analyses, which minimized the risk of interference between participant responses in each administration.

In addition to the threat of interference of other mathematics professional development programs, each teacher had a different exposure to mathematics content experts within their school setting. Some schools had teachers within certain grade levels who are math facilitators for the MSP grant project, and thus experts, in the *Investigations* curriculum. Therefore, each teacher’s exposure to mathematics content, mathematics teaching practices, and content experts varied. This threat to internal validity is difficult to minimize, but the assumption of the researcher is that the variance within the participant population will minimize the effects of individual teacher's experiences.

SETMI responses for this study were not used to make predictions, which minimized the overall impact of the professional development aspect of the MSP grant as interfering with the evidence of validity of the SETMI. Instrumentation as a threat to internal validity was minimized by ensuring that the SETMI meets content and face validity before further analysis was conducted. The examination of the revised SETMI items by a focus group ensured that instrumentation was not a threat to the internal validity of this study.

Finally, attrition in this study was minimized by the context of the participants being part of a larger grant project for which they chose to apply to be a participant. Although some Cohort II participants were lost due to attrition in the year-long period of the MSP grant project (n = 183 for pre-test and n = 153 for post-test), the data analysis conducted in this study is not affected by loss of those participants. In some cases, such as the examination of concurrent validity, the attrition of teachers or the presence of missing data may have caused a slight difference in the number of participants for that data analysis. However, these differences are slight and are not thought to affect the integrity of that data analysis.

Minimizing Threats to External Validity

It was important to minimize threats to external validity of this study as the objective of this study was to create a reliable and valid instrument to be used with elementary teachers. One threat to the generalizability of this study was the participant group as a representative sample of the elementary teacher population. Teachers in the participant sample were mostly Caucasian and mostly female, which echoed the population of elementary school teachers. Participants were also evenly distributed among grade levels from grades K to 5. Although this threat is minimized by the demographics of the participant sample as being representative of the population of elementary school teachers, it was possible that teachers who chose to apply to the MSP grant either had a high self-efficacy for mathematics or a low self-efficacy for mathematics, which impacted their decision to apply. In District One, teachers were chosen for participation based on their application. In District Two, teachers were chosen to participate in this study in an effort to have all elementary teachers participate in either the MSP grant project or another mathematics professional development program. To some degree, the range of experience and self-efficacy with mathematics in the participant group should account for individual differences.

The same concerns could cause selection-treatment interaction to be a threat to external validity; however the diversity in the group of participants and the large number of participants in this study over the two years minimized that threat. Specificity of variables is minimized as a threat to external validity by establishing face and content validity of the SETMI as representing all aspects of elementary mathematics teaching. Therefore, variables are not overly specific to any one sub-group of elementary teachers (i.e. grade or district). Experimenter effects are minimized because the researcher in this study was not a part of the professional development team and serves only as an external evaluator on the MSP grant project. Finally, participant effects were minimized because participants understand that completing the SETMI was a part of their MSP grant responsibilities. An effort was also made to spread teacher assessments and surveys out over time as to avoid giving teachers too many instruments to complete at once as a part of their grant responsibilities.

Chapter Summary

The purpose of this dissertation study was to examine the validity of a newly created instrument to measure self-efficacy of elementary mathematics teachers; the Self-efficacy for Teaching Mathematics Instrument (SETMI). Although in the past 30 years teacher self-efficacy has been studied diligently, there exist few instruments for measuring mathematics teacher self-efficacy. To date, the most widely used instrument for the purpose of measuring general teacher self-efficacy is the TSES created by Tschannen-Moran and Woolfolk Hoy (2001). Pajares (1997) and Bandura (1994) both assert that the most accurate measures of self-efficacy are one's self-assessments of their ability to complete relevant and specific sub-tasks of a task. If teaching mathematics is the task where self-efficacy is to be measured, specific sub-tasks associated with teaching mathematics should be included on any instrument to measure teacher self-efficacy for teaching mathematics. Other content-specific instruments that have been created fail to ask specific questions.

In an effort to create an instrument to measure the self-efficacy of elementary mathematics teachers, three facets of teaching elementary mathematics were considered (Ernest, 1989). First, the teachers' overall ability to create mathematics lessons and individualize mathematics instruction should be assessed. Second, the teacher's own mathematics self-efficacy should be assessed by asking questions about their ability to understand specific mathematics content. Finally, the teacher's mathematics teaching self-efficacy should be assessed by asking the teacher to rate their ability to teach that mathematics content. If all three facets of a teacher's complex self-efficacy for teaching elementary mathematics could be assessed, a greater understanding of a teacher's belief district could be gained thus allowing for personalized professional development experiences.

Creating an instrument to measure self-efficacy is a lengthy and complex task. An instrument must meet rigorous guidelines for reliability and validity. In order to examine reliability of the instrument, Cronbach's alpha was used. Test-retest reliability was also examined to ensure that participant's scores remain stable over time. In order to examine evidence for validity of the SETMI, several procedures were performed. Content and face validity were met by allowing experts in self-efficacy beliefs and elementary mathematics education and a focus group of elementary mathematics teachers to suggest revisions to the instrument and approve the content items. Concurrent validity of the SETMI was explored by correlating Part Two and Part Three of the SETMI with Part One of the SETMI. Part One of the SETMI uses items from the TSES, a valid measure of teacher sell-efficacy, and is thus a measure of a similar construct. Finally, construct validity was explored by using factor-level analysis (EFA and CFA) to examine the factor structure of the observed variables and confirm the theoretical construct structure.

CHAPTER 4: RESULTS

The purpose of this study was to show evidence of reliability and validity for a survey instrument that was created to measure the mathematics self-efficacy of elementary teachers. The SETMI was administered to two participant groups as a part of the external evaluation of an MSP grant project. Statistical procedures were performed on the instrument as it was under development. Those procedures are highlighted in this chapter as evidence of reliability and validity of the instrument. This chapter also describes the factor structure and the underlying constructs of the SETMI in an effort to establish reliability and validity evidence for the instrument. This chapter will also highlight the findings relevant to the theoretical model and model fit.

Brief Summary of SETMI Development

Based on a review of the literature, the SETMI was developed using two instruments as a foundation, along with Bandura's ideas about self-efficacy from Social Cognitive Theory, literature about mathematics beliefs (Ernest, 1989), and input from content experts. Items from the Teacher's Sense of Efficacy Scale (TSES) (Tschannen Moran & Woolfolk Hoy, 2001) and the Teaching Mathematics in Inclusive Settings instrument (Aerni, 2008) were used, along with the researcher's creation of items, to develop the SETMI in July 2010.

For the SETMI (Version One), there were three expected constructs prior to further analysis of data: *mathematics teaching self-efficacy* (MTSE) *(items 1-9)*, *efficacy for understanding elementary mathematics content* (EUMC) (items 10 through 24), and *efficacy for teaching mathematics content* (ETMC) (items 25 through 39).

Reliability

Data collection from Cohort II of the MSP grant (see Chapter 3) was used to establish reliability of the SETMI. One hundred eighty three instruments were distributed and completed. Reliability was explored using Cronbach's alpha. For MTSE, the reliability is .88, for EUMC, the reliability is .79, and for ETMC, the reliability is .91. These are acceptable reliability values and indicate moderately strong reliability of the instrument (Cronbach & Meehl, 1955). Because the reliability of the instrument was deemed to be acceptable, further examination of the factor structure was conducted.

The SETMI was modified based on results of an EFA and feedback from mathematics content experts. This process is detailed below. For Version Two of the instrument, the new reliabilities for each factor were: MTSE (.85), EUMC (.83), and ETMC (.93). This is much improved from Version One and shows the improved stability of the instrument. Table 3 shows the increase in reliability from Version One to Version Two. Table 3 also displays the Cronbach's alpha for both Version One and Version Two of the SETMI along with means and standard deviations for each construct.

Table 3.

Reliability and Descriptive Statistics of the Constructs (pre, N = 183; post, N = 153)

	Version One			Version Two		
	α	*M*	*SD*	*α*	*M*	*SD*
MTSE	.88	6.24	1.08	.85	3.72	.56
EUMC	.79	3.71	.55	.83	3.33	.65
ETMC	.91	3.65	.58	.93	3.58	.70

Note. Version One of the SETMI had a rating scale from 1 to 9 whereas the rating scale of Version Two was 1 to 5.

Factor Structure of the SETMI (Version One)

Due to the construction of the SETMI as containing three distinct parts, three separate factor analyses were conducted using SPSS version 15. It was determined *a priori* that any item with a factor loading greater than .40 would be considered. This is based on the suggestion of Comrey and Lee (1992) that loadings in excess of .71 are considered excellent, .63 very good, .55 good, .45 fair, and .32 poor. Loadings less than .32 (less than 10% overlapping variance) are usually not interpreted (Comrey & Lee, 1992).

Exploratory Factor Analysis (EFA) was used to determine the factor structure of the data. Principal Component Analysis and Varimax rotation were used. Scree plots were also examined. The number of factors was not set before the EFA was conducted. Two principal components emerged from Part One of the SETMI, explaining 67.54% of the total variance. However, examination of the component matrix revealed that only one item (Item 8) was in the second component. The low correlation between Item 8 and other items in this factor indicated that it was not a valid item for self-efficacy and therefore Item 8 was not used in further analysis. Item correlations can be found in Table 4. Hence, data indicate that only one factor exists in Part One of the SETMI; *mathematics teaching self-efficacy (MTSE).*

Table 4.

Item Correlations, SETMI Version One, Part One (MTSE) (n = 183)

	1	2	3	4	5	6	7	8	9
1	--	.77**	.43**	.63**	.44**	.43**	.45**	.23**	.55**
2		--	.58**	.68**	.47**	.42**	.48**	.25**	.53**
3			--	.57**	.50**	.55**	.52**	.32**	.57**
4				--	.53**	.49**	.56**	.23**	.50**
5					--	.58**	.61**	.10**	.53**
6						--	.73**	.11**	.62**
7							--	.15**	.65**
8								--	.32**
9									--

Note. ** $p < .01$.

The second EFA revealed three factors in Part Two of the SETMI: (a) *understanding mathematics concepts* (items 10, 11, 12, 14, 15, 17, 18); (b) *analyzing mathematics problems* (items 19, 20, 21, 22, 24); and (c) *solving multi-step mathematics problems* (items 13, 16, 23). These three factors explained 60.80% of the total variance. Item 23 was cross-loaded into both factor 2 (.54) and factor 3 (.60). Most items in Part Two were statistically significantly correlated (see Table 5) and therefore these items were grouped into the second-order factor *efficacy for understanding mathematics content* (EUMC) for further analysis.

Table 5.

Item Correlations, SETMI Version One, Part Two (EUMC) (n = 183)

	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
10	--	.59**	.61**	.31**	.43**	.38**	.21**	.34**	.40**	.27**	.13	.26**	-.01	.13	.41**
11		--	.78**	.34**	.60**	.45**	.20**	.42**	.55**	.34**	.21**	.41**	.10	.09	.44**
12			--	.39**	.59**	.45**	.27**	.43**	.55**	.30**	.09	.36**	.08	.17*	.39**
13				--	.48**	.31**	.57**	.31**	.37**	.29**	.10	.27**	.26**	.47**	.16*
14					--	.56**	.39**	.39**	.55**	.40**	.15	.46**	.19*	.23**	.37**
15						--	.35**	.49**	.45**	.42**	.15*	.27**	.08	.14	.26**
16							--	.38**	.34**	.25**	.09	.26**	.23**	.43**	.08
17								--	.56**	.40**	.12	.20**	.17*	.20*	.29**
18									--	.52**	.20**	.43**	.18*	.29**	.47**
19										--	.27**	.47**	.34**	.30**	.40**
20											--	.36**	.30**	.32**	.33**
21												--	.48**	.31**	.51**
22													--	.47**	.41**
23														--	.37**
24															--

Note. *$p < .05$; ** $p < .01$.

The third EFA revealed three factors in Part Three of the SETMI that explained 64.83% of the total variance: (a) *teaching students mathematics concepts* (items 25, 26, 27, 29, 32, 33); (b) *teaching students to analyze mathematics problems* (items 34, 35, 36, 37, 38, 39); and (c) *teaching students to solve multi-step mathematics problems* (items 28, 30, 31). Item 38 was cross-loaded into both factor 2 (.64) and factor 3 (.50). Most items in Part Three were statistically significantly correlated (see Table 6) and therefore grouped into the second-order factor *efficacy for teaching mathematics concepts* (ETMC) for further analysis.

Table 6.

Item Correlations, SETMI Version One, Part Three (ETMC) (n = 183)

	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
25	--	.67**	.64**	.24**	.41**	.33**	.16**	.43**	.47**	.33**	.17*	.22**	.11	.11	.42**
26		--	.82**	.37**	.52**	.38**	.27**	.48**	.55**	.38**	.25**	.34**	.20*	.24**	.40**
27			--	.36**	.63**	.35**	.21**	.43**	.52**	.32**	.19*	.28**	.08	.26**	.36**
28				--	.58**	.38**	.57**	.39**	.28**	.34**	.38**	.50**	.41**	.61**	.37**
29					--	.50**	.44**	.44**	.49**	.35**	.35**	.46**	.22**	.38**	.39**
30						--	.47**	.47**	.31**	.42**	.21*	.37**	.26**	.26**	.32**
31							--	.39**	.29**	.49**	.30**	.51**	.53**	.53**	.25**
32								--	.49**	.50**	.31**	.34**	.29**	.31**	.36**
33									--	.50**	.33**	.38**	.30**	.29**	.51**
34										--	.35**	.51**	.45**	.42**	.43**
35											--	.54**	.38**	.37**	.40**
36												--	.61**	.58**	.46**
37													--	.59**	.42**
38														--	.44**
39															--

Note. *p < .05; **p < .01.

Concurrent validity of the SETMI (Version One)

Concurrent validity of the SETMI (Version One) was examined after conducting the EFA. The purpose of this analysis was to provide evidence that items in Part Two and Part Three of the SETMI, the parts created solely by the researcher, were accurate measures of self-efficacy. Part One of the SETMI was compared against Part Two and Part Three. Part One uses items of the TSES re-worded to be mathematics specific and since the TSES has been validated, Part One represents a validated instrument. A mean score for MTSE, EUMC, and ETMC was computed for each participant. Correlations between each of the factors (MTSE, EUMC, and ETMC) were examined and all were statistically significantly different from zero at the .01 level: MTSE and EUMC (r = .34); and MTSE and ETMC (r = .32). Although the correlations are not incredibly high, their status as statistically significant confirms the concurrent validity of the instrument. These results are located in Table 7.

Table 7.

Version One Factor Correlations and Concurrent Validity (n = 183)

	MTSE	F2	F3	F4	F5	F6	F7	EUMC	ETMC
MTSE	--	.29**	.29**	.25**	.31**	.28**	.17*	.34**	.32**
F2		--	.53**	.46**	.85**	.37**	.51**	.92**	.71**
F3			--	.44**	.55**	.78**	.48**	.78**	.73**
F4				--	.37**	.42**	.61**	.65**	.51**
F5					--	.56**	.56**	.82**	.89**
F6						--	.63**	.60**	.86**
F7							--	.62**	.78**
F8								--	.82**
F9									--
Mean	6.37	3.46	3.74	4.33	3.27	3.82	4.09	3.73	3.65
SD	1.15	.75	.59	.55	.78	.62	.64	.55	.59

Note. (a) Factor 1 is Mathematics Teaching Self-efficacy, Factor 2 is Understanding Math Concepts, Factor 3 is Analyzing Math Concepts, Factor 4 is Solving Multi-step Problems, Factor 5 is Teaching Math Concepts, Factor 6 is Teaching Students to Analyze Math Concepts, Factor 7 is Teaching Students to Solve Multi-Step Problems; (b) Factor 1 is measured on a 9-point scale while all other factors are measured on a 5-point scale; (c) ** $p < .01$, * $p < .05$.

Table 7 also shows correlations of each of the sub-factors identified on the EFA. A mean score for each item in the sub-factor was computed and this was used as a score for each participant on the sub-factor. Having sub-factors for each of the factors in the model does not align with the original theoretical framework of this study. Therefore, the sub-factor scores were not used in further analysis.

Correlation between Pre and Post Administration

Cohort II participants completed the SETMI (Version One) as prescribed in August 2010 and February 2011. For the purposes of the MSP grant, the August administration served as a pre-test and the February administration served as a post-test. Participants had participated in at least 60 hours of professional development in mathematics between the pre and post-test administrations.

The correlation between pre to post administration was calculated by correlating the pre-test scores on the three constructs MTSE, EUMC, ETMC to the post-test scores for the same three constructs. The correlations between pre and post scores were: for MTSE, r = .39; for EUMC, r = .56; and for ETMC, r = .41. It should be noted that there was some attrition between the pre-test (n = 183) and post-test (n = 153) administration, but all correlations were significant at the .01 level. The attrition could be caused by the administration of the SETMI during one of the follow-up workshops for which some teachers were not excused from their teaching duties to attend. Other causes for missing data from pre to post-test could be attrition from the MSP grant project. Efforts were made to collect these surveys from teachers who were not present at the workshop by allowing for online submissions. For a summary please reference Table 8.

Table 8.

Correlation between Pre and Post Administration (pre, N = 183; post, N = 153)

	MTSE_Post	EUMC_Post	ETMC_Post
MTSE_Pre	.39**	--	--
EUMC_Pre	--	.56**	--
ETMC_Pre	--	--	.41**

Note. ***p* < .01.

Revisions to the SETMI (Version One)

The results of the three Exploratory Factor Analyses and concurrent validity indicated that the SETMI needed revisions. First, the factor structure of the SETMI was more complex than initially intended. The theoretical framework of this study indicated that there are three constructs of self-efficacy for teaching mathematics in elementary school: *mathematics teaching self-efficacy* (MTSE), *efficacy for understanding mathematics content* (EUMC), and *efficacy for teaching mathematics content* (ETMC). The cross-loadings of some items into more than one factor on the EFA indicated that some items are not clear, observable correlates to these factors. Upon closer review, some items needed revision to be more specific and measurable. Additionally, the small number of items in some of the factors indicated the need for more items to be created to capture that construct.

In response to these issues, the researcher and primary content expert, Dr. Drew Polly, examined the items closely being attentive to their complexity and their relevance to the state standard course of study for Kindergarten through fifth grade. An attempt was made to equally represent items from each of the grade levels. Mathematics concepts build on each other through the grades, so this was thought to be appropriate. Additionally, elementary teachers are licensed in this state as being qualified to teach Kindergarten through sixth grade. This license essentially means that an elementary teacher should be able to teach mathematics content in any of the grades within the elementary spectrum.

In order to have an instrument that is more closely aligned to the mathematics content that this population teaches and the theoretical model, the following changes were made. Item 8 loaded into a second factor during the EFA and was therefore removed from further inclusion in analyses as mentioned previously and deleted from Version Two of the SETMI. Item 9 was added to the instrument by the content experts during the development of the SETMI, but it was removed because further analysis suggested it may not be a clear indicator of self-efficacy.

The rating scale for Part One was changed to reflect the five-point scale used in Parts Two and Three. Bandura (2006) touted the use of broad scales on instruments of this nature in order to capture the variance in responses. More recent efforts have shown that a five-point scale, or even a three-point scale, is adequate (Dawes 2002; 2008). The wording on the five-point scale was also changed to reflect the wording on the original TSES scale because it fit the wording of the items better and aligned with the theory of self-efficacy as determining how well someone can perform an action to create a result. The original wording on Version One was “Not Confident, Barely Confident, Somewhat Confident, Confident, and Very Confident.” The wording for Version Two, on all three parts is now “None at All, Very Little, Strong Degree, Quite a Bit, and A Great Deal.” This is more appropriate as the latter is a rating of one’s ability to perform an action, not ones confidence in their ability to perform an action.

For Part Two and Three of the SETMI, the wording was changed from “How confident do you feel in your understanding of” to “How well do you understand.” This was in an effort to better align with both Bandura’s (1996) and Pajares’ (1997) thoughts that self-efficacy item-wording should be “will or can” instead of a measure of confidence. This also aligns with Dellinger et al.’s (2008) findings that the wording “My belief in my ability to…” had different results when comparing responses to items worded “I can” or “I am able to.” The wording for Part Three was changed from “How confident do you feel in your ability to teach students to” to “How well can you teach students to” for the reasons stated above.

Other items in Part Two and Three were simplified so that items only measure one specific aspect of mathematics content. For a complete view of all edits to mathematics content items please see Tables 9 and 10.

Table 9.

Item Revisions from SETMI Version One to SETMI Version Two (Part Two Only)

#	Version One Item	#	Version Two Item
10.	Integers, rational, and irrational numbers.	8.	Characteristics of Numbers (i.e. whole numbers, rational/ irrational numbers).
11.	The equivalence of fractions, decimals, and percents.	9.	Strategies for composing and decomposing numbers by manipulating place value in addition and subtraction.
		10.	Strategies for composing and decomposing numbers by manipulating place value in multiplication and division.
12.	Arithmetic operations on decimals and fractions.	11.	Conversion of a fraction to a decimal and vice versa.
		12.	Comparing equivalence of fractions and decimals.
~~13.~~	~~Solving one to two step arithmetic word problems.~~		
14.	Inverse relationships between multiplication and division.	13.	Inverse relationships between operations (i.e. +, - and *, ÷).
15.	Coordinate planes.	14.	Coordinate planes.
16.	Interpreting bar and line graphs.	15.	Collecting, plotting and interpreting data (on any type of graph).
~~17.~~	~~Use of compasses, rulers, and protractors.~~		
18.	Square and cubic units.	16.	Measurement of area and perimeter.
19.	Size, quantity, and capacity.	17.	Conversions between units in the same system (i.e. grams → kilograms, inches → yards).
		18.	Conversions between units in a different system (i.e. kilograms → pounds, inches → centimeters).
~~20.~~	~~Use of websites to promote mathematical understanding.~~		
21.	Use of estimation as a problem-solving strategy.	22.	Probability of outcomes.
22.	Identifying, describing, and creating patterns.	20.	Mathematical patterns.
~~23.~~	~~Solving one and two-step equations.~~		
~~24.~~	~~Different representations to describe a proportional relationship.~~		
		19.	Measuring the length of objects.
		21.	Variables in an algebraic equation.

Table 10.

Item Revisions from SETMI Version One to SETMI Version Two (Part Three Only)

#	Version One Item	#	Version Two Item
25.	Understand integers, rational, and irrational numbers.	23.	Describe characteristics of Numbers (i.e. whole numbers, rational/ irrational numbers).
26.	Describe equivalence of fractions, decimals, and percents.	24.	Perform strategies for composing and decomposing numbers by manipulating place value in addition and subtraction.
		25.	Perform strategies for composing and decomposing numbers by manipulating place value in multiplication and division.
27.	Perform arithmetic operations on decimals and fractions.	26.	Convert a fraction to a decimal and vice versa.
		27.	Compare equivalence of fractions and decimals.
~~28.~~	~~Solve one to two step arithmetic word problems.~~		
29.	Understand inverse relationships between multiplication and division.	28.	Interpret inverse relationships between operations (i.e. +, - and *, ÷).
30.	Locate points on a coordinate plane.	29.	Manipulate coordinate planes.
31.	Interpret bar and line graphs.	30.	Collect, plot, and interpret data (on any type of graph).
~~32.~~	~~Use compasses, rulers, and protractors.~~		
33.	Understand square and cubic units.	31.	Measure area and perimeter.
34.	Measure size, quantity, and capacity.	32.	Convert between units in the same system (i.e. grams → kilograms, inches → yards).
		33.	Convert between units in a different system (i.e. kilograms → pounds, inches → centimeters).
~~35.~~	~~Use of websites to promote mathematical understanding.~~		
36.	Use estimation as a problem-solving strategy.	27.	Interpret probability of outcomes.
37.	Identify, describe, and create patterns.	35.	Discover and create mathematical patterns.
~~38.~~	~~Solve one and two-step equations.~~		
~~39.~~	~~Use different representations to describe a proportional relationship.~~		
		34.	Measure the length of objects.
		36.	Interpret variables in an algebraic equation.

Some items from Version One were deleted for various reasons. Item 13, and therefore item 28, ("solving one to two step arithmetic word problems") originally came from Aerni's (2008) instrument. It loaded into a sub-factor of Part Two and did not provide a concise way to measure the complexity of solving word problems. Those items were deleted and better worded items to describe process of solving mathematical equations were added (e.g.,"inverse relationships between operations," "variables in a mathematical equation," and "characteristics of numbers"). Item 20 ("use of websites to promote mathematical understanding"), and therefore item 35, were also deleted as they do not relate specifically to the standard course of study for grades Kindergarten through fifth grade. This item had a low correlation between other items in the same construct.

One final addition to Version Two of the SETMI is an indicator where teachers can choose the grade level(s) that they have taught. The reason for this addition comes from Tschannen Moran and Woolfolk Hoy's (2001) thoughts that self-efficacy can vary according to context. Teaching lower or upper elementary grades could have an impact on a teacher's self-efficacy for mathematics content and is worthy of further analysis in future studies.

As stated in Chapter 3, a focus group of teacher-leaders examined the SETMI Version Two in the summer of 2011. The only suggestion for revisions from the focus group was to add aspects of probability to the instrument. Item 23 (Version Two) now states, "Probability of outcomes." The focus group also suggested that Item 17 only state, "Measurement of area and perimeter" instead of the original "Measurement of area and perimeter of rectangles." Otherwise the focus group felt that the newly revised items were more clear and relevant to the elementary state standard course of study. Version Two of the SETMI now contains 37 concise items. The first seven items are indicators of MTSE, items 8 through 22 are EUMC, and items 23 through 37 are ETMC. Each of these constructs is subsumed over the second-order construct of *self-efficacy for teaching mathematics* (SETM).

Model Fit of the SETMI

In order to verify the theoretical model proposed in this study, that self-efficacy for mathematics contains three distinct factors: MTSE, EUMC, and ETMC, a Confirmatory Factor Analysis (CFA) was conducted using LISREL. First, imputation of missing values was calculated in the data set using the mean of responses within the construct for each participant. Then, using syntax in LISREL, the model fit was tested against raw data from the August 2011 administration of the SETMI with Cohort III.

The first model tested (Self-efficacy Model 1) was the original theoretical model containing three latent variables within the second-order construct of *self-efficacy for teaching mathematics* (SETMI). This model was found not to fit the covariance matrix (χ^2 = 2562.17, df = 626, p < .0001; RMSEA [90%CI] = .12, .13; CFI = .90; GFI = .59). Pattern coefficients were within reason, but some error variances of observed variables were quite large (e.g., Item 17 = .90, Item 34 = .73). The model was tested with Item 17 eliminated, but this did not improve the model fit. Additionally, the correlation between the second-order construct and the first-order constructs was quite high for EUMC (.82) and ETMC (.99). This model can be seen in Figure 1.

The modification indices suggested allowing error variances between observed variables to correlate but, suggested modifications were not allowed because they did not fit the theoretical framework of the study. As this model was not a good fit to the data, other models were considered.

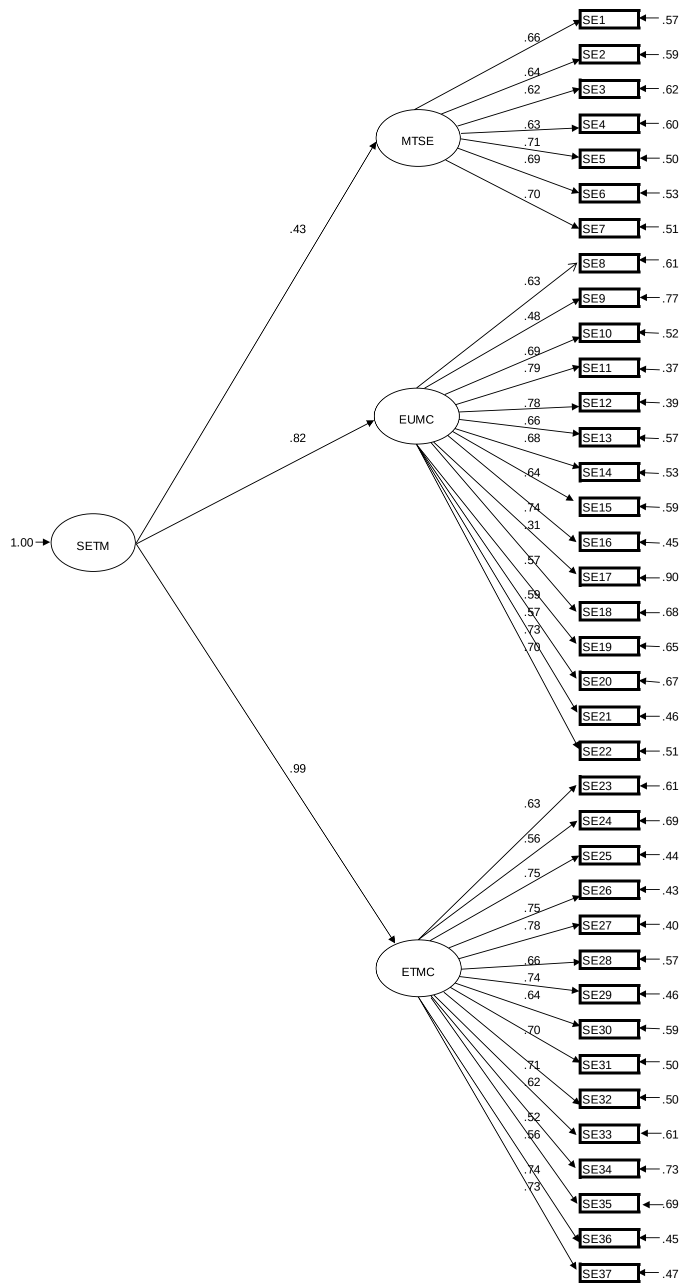

Figure 1. Original *self-efficacy for teaching mathematics* (SETM) model with three latent constructs*: mathematics teaching self-efficacy* (MTSE), *efficacy for understanding mathematics content* (EUMC), and *efficacy for teaching mathematics content* (ETMC). Observed variables in this model come from the Self-efficacy for Teaching Mathematics Instrument (SETMI) version two that was developed and revised as a part of this study.

Modified Model Fit

Upon considering other options for the model, the construct EUMC was examined closely as some of the items within that construct were problematic within the model fit (Item 17 and Item 9). Upon closer review of item wording it was determined that Items 8-22 were possibly not measures of self-efficacy for content but were perhaps a self-report of mathematics content knowledge. Model 2 for self-efficacy was considered without these items, although these items could relate to self-efficacy as they measure self-concept of one's mathematics content knowledge.

The items in ETMC were closely examined and two logical testlets were formed based on item content in that construct: *efficacy for teaching mathematics operations* (ETMO) and *efficacy for teaching mathematics concepts* (ETMC). Testlets are groups of questions that are related to larger stimuli and administered together. Testlets normally contain four or more items that are always presented in the same order. The benefit of the testlet is that it makes testing more efficient and allows for greater control in examining constructs within a test (Wainer, Bradlow, & Wang, 2007). ETMO consisted of items 24, 25, 26, 27, 31, 32, 33, and 34. All of these items were relevant to teaching someone to perform a mathematical procedure. All items in this testlet are statistically significantly correlated (see Table 11).

Table 11.

Item Correlations for Testlet Efficacy for Teaching Mathematical Operations (ETMO)

	24	25	26	27	31	32	33	34
24	--	.58**	.33**	.40**	.40**	.28**	.23*	.36**
25		--	.58**	.63**	.50**	.53**	.47**	.35**
26			--	.84**	.51**	.55**	.50**	.30**
27				--	.57**	.59**	.55**	.32**
31					--	.49**	.41**	.43**
32						--	.79**	.37**
33							--	.26**
34								--

Note. ** $p < .01$.

ETMC consisted of items 23, 28, 29, 30, 35, 36, and 37. All items were relevant to teaching someone a mathematical concept or idea, but not a procedure. All items were statistically significantly correlated (see Table 12).

Table 12.

Item Correlations for Testlet Efficacy for Teaching Mathematics Concepts (ETMC)

	23	28	29	30	35	36	37
23	--	.47**	.39**	.40**	.43**	.39**	.36*
28		--	.61**	.46**	.41**	.48**	.47**
29			--	.58**	.36**	.59**	.55**
30				--	.51**	.44**	.56**
35					--	.44**	.56**
36						--	.65**
37							--

Note. ** p < .01.

Both parts of mathematics are important and complimentary to one another. Teaching procedures is different from teaching concepts as one is more concrete and the other abstract. It is the researcher's belief that different judgments of ability may influence the self-efficacy of teachers to perform both teaching tasks. This is aligned to Peterson, Fennema, Carpenter, and Loef's (1989) conceptual framework for examining teacher's pedagogical content beliefs in mathematics. One construct they explored was the idea that mathematics skills should be taught in relation to understanding and problem solving.

Self-efficacy Model 2 had three latent constructs for Self-efficacy for Mathematics Teaching. The three latent constructs were: MTSE, ETMO, and ETMC. This model had better fit to the covariance matrix (χ^2 = 566.86, *df* = 205, p < .0001;

RMSEA [90%CI] = .09, .10; CFI = .95; GFI = .80). Pattern coefficients were within reason and error variances of observed variables were acceptable as most were around .50 (Schumaker & Lomax, 2010). The modification indices suggested allowing several error covariances to correlate, most were within reason as they maintained the theoretical framework. The error covariance was allowed to correlate between Item 26 and Item 27, as both questions regarded mathematical procedures for fractions and decimals. Error covariance was also allowed to correlate between Item 32 and Item 33, as both items concerned conversion of measurement units. Additionally, there was a high correlation between ETMO and ETMC (.95) and this was expected due to the nature of these items as concerning self-efficacy for teaching specific mathematical content. This model can be seen in Figure 2. Table 13 shows the change in model fit between Self-efficacy Model 1 and Self-efficacy Model 2.

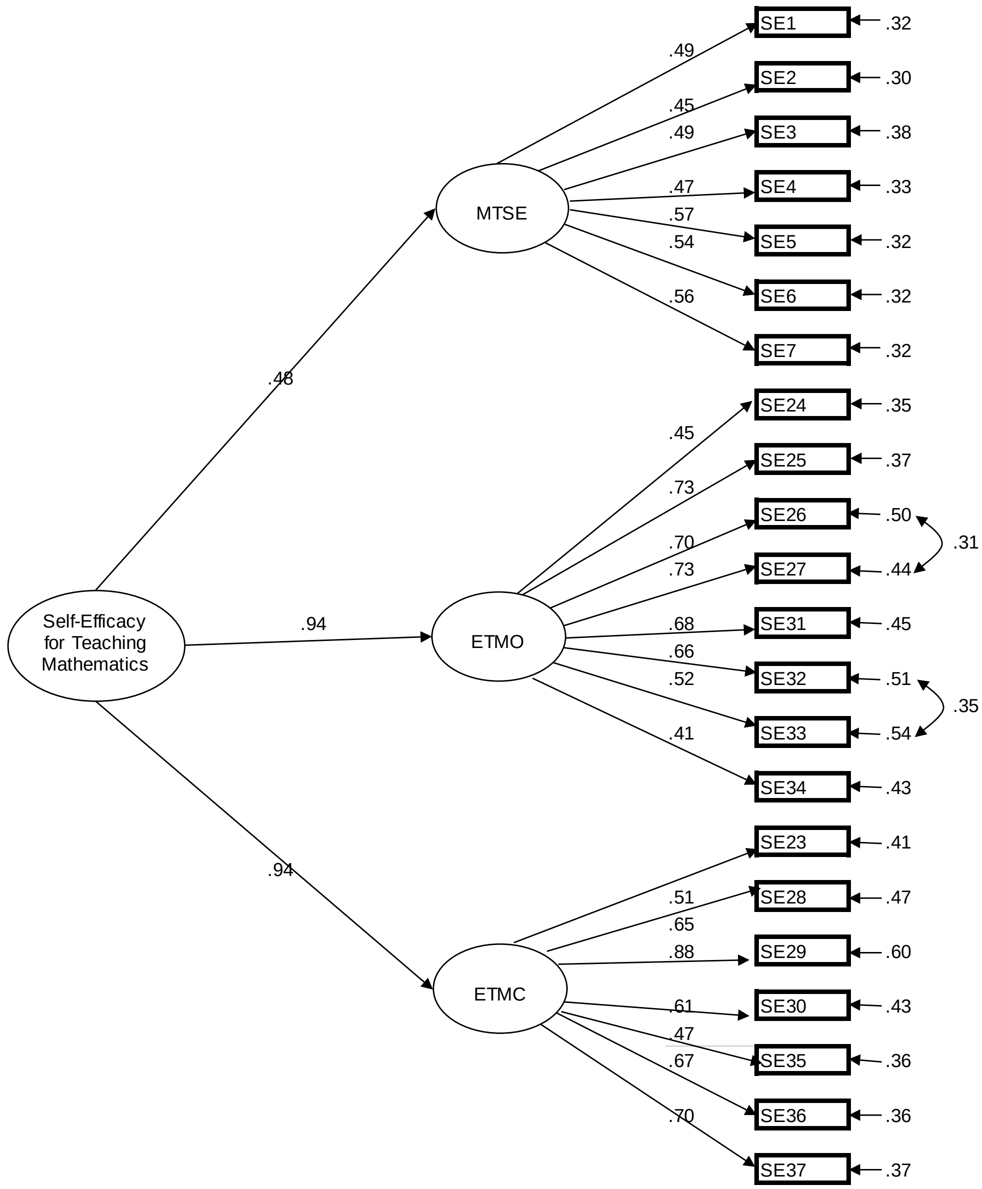

Figure 2. Modified model with three latent constructs*: mathematics teaching self-efficacy* (MTSE), *efficacy for teaching mathematics operations* (ETMO), and *efficacy for teaching mathematics concepts* (ETMC). Observed variables in this model come from the Self-efficacy for Teaching Mathematics Instrument (SETMI) version two that was developed and revised as a part of this study.

Table 13.

Model Fit Indices for the SETMI

	χ^2	*df*	GFI	AGFI	NFI	NNFI	CFI	SRMR	RMSEA	90% LL	90% UL
Model 1	2562.17	626	.59	.54	.87	.89	.90	.083	.13	.12	.13
Model 2	566.86	205	.80	.75	.92	.94	.95	.073	.09	.09	.10
Model 3	448.19	125	.81	.74	.91	.92	.94	.086	.11	.09	.12

Note . GFI=Goodness of Fit, AGFI=Adjusted Goodness of Fit, NFI=Normed Fit Index, NNFI=Non-Normed Fit Index, CFI=Comparative Fit Index, SRMR=Standardized Root Mean Residual, RMSEA=Root Mean Square Error of Approximation.

In a further attempt to better understand the complexity of an elementary teacher's mathematics self-efficacy, an multivariate analysis of variance (MANOVA) was conducted to examine the current grade level of teachers and their scores on each of the original constructs (MTSE, EUMC, ETMC). SPSS version 19 was used. The dependent variable was scores for each of the three constructs and the independent variable was grade level. The assumption of equality of covariance matrices was not satisfied (Box's $M = 103.45$, $p < .01$). There was a significant difference between the treatment (grade level) on the combined dependent variables (MTSE, EUMC, ETMC), *Wilks' Lambda* = .71, $F = 3.67$, $p < .01$. Univariate tests were performed to examine the differences between the groups on the dependent variable, grade level. There was a difference between the groups on EUMC ($p < .01$) and for ETMC (p < .01) but not for MTSE ($p = .12$).

As MTSE was not significantly different for grade levels, and as EUMC is potentially not a good measure of self-efficacy, an analysis of variance (ANOVA) was conducted with ETMC as the dependent variable and grade level as the independent variable. There was a statistically significant difference ($p < .01$) between the groups with regard to ETMC. A boxplot shows (Figure 3) that only 5th grade was normally distributed, with all other grades (K, 1, 2, 3, 4) being positively skewed.

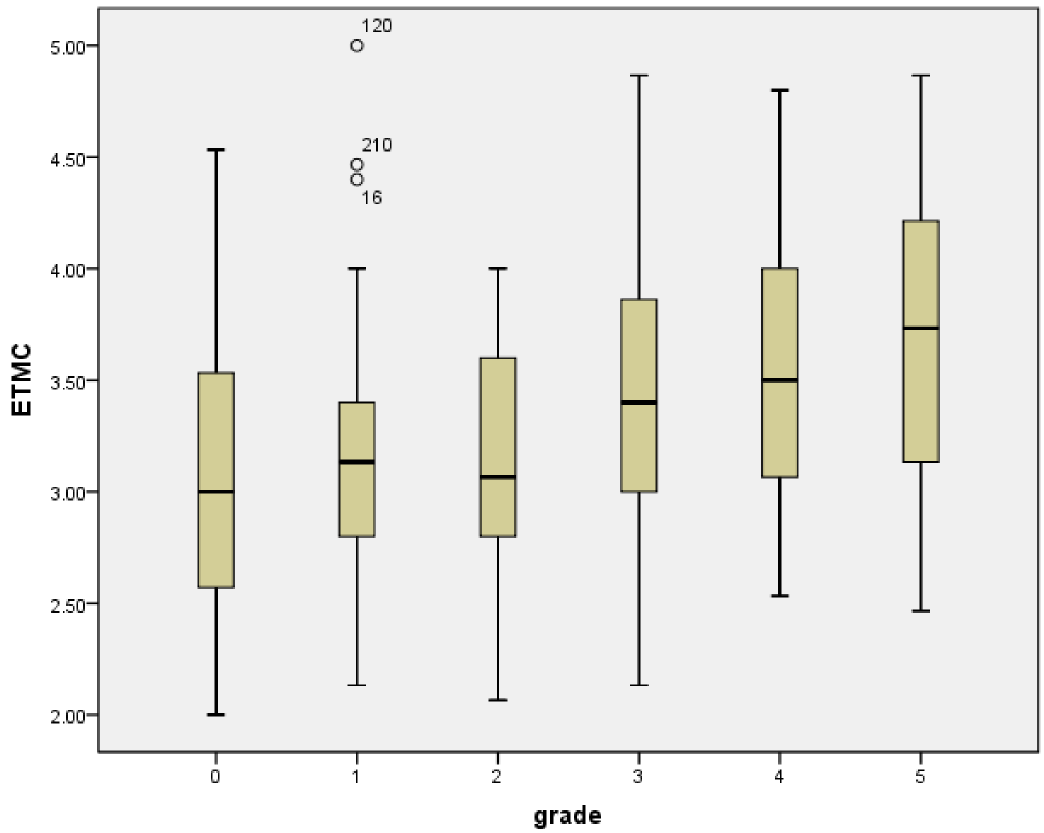

Figure 3. Box plot of ETMC by grade level of participants ($n = 183$).

Kindergarten showed the most extreme positive skew. Kindergarten was statistically significantly different from 3rd ($p = .04$), 4th ($p = .001$), and 5th grade ($p = .001$). First grade was statistically significantly different from 4th ($p = .02$) and 5th grade ($p = .01$). Second grade was statistically significant from fifth grade ($p = .04$). As the mathematics content covered in a Kindergarten class is much more basic than other grades, and as it is statistically significantly different from upper elementary grades, Kindergarten teachers were withdrawn from further analysis of the data. The certification of the Kindergarten teachers in this dataset is also unknown. In the state where this study takes place, Kindergarten teachers can either be certified “Birth through Kindergarten” or “Kindergarten through Sixth Grade.” The academic content presented in teacher preparation programs for these certifications differs greatly. A teacher who is certified “Kindergarten through Sixth Grade” would likely have more mathematics courses in teacher preparation than a teacher certified as “Birth through Kindergarten” (K. Lynch-Davis, personal communication, February 16, 2012). Curriculum for teacher preparation of Birth through Kindergarten teachers focuses mainly on developmentally appropriate behaviors and not academic content (D. Brewer, personal communication, February 17, 2012).

An EFA was conducted on the responses for ETMC only to determine if removal of Kindergarten from the dataset altered the construction of the factor, thus allowing for smaller testlets to be formed. The EFA revealed three factors that explained 66.12% of the variance. Factor loadings higher than .40 were considered. The three factors are: *numbers and operations* (NO) (items 24, 25, 23, and 28), *transformation of numbers* (TN) (items 26, 27, 32, 33), and *interpretation of data* (ID) (items 34, 30, 35, and 37). Three items cross-loaded (Item 29, 31, and 36) and those items were not included in further analysis.

Three measurement models, one for each of the three new constructs (NO, TN, and ID) were tested using LISREL. The first measurement model for NO fit the covariance matrix (χ^2 = 18.29, df = 2, p < .0001; RMSEA [90%CI] = .13, .31; CFI = .95;

GFI = .95).However, the model was still statistically significant. Modification indices suggested allowing error variance to correlate between Item 23 and 24. This improved the model fit (χ^2 = 1.35, *df* = 1, *p* = .24; RMSEA [90%CI] = .00, .21; CFI = 1.00; GFI = 1.00). The second measurement model for TN did not fit the covariance matrix (χ^2 = 28.21, *df* = 2, *p* < .0001; RMSEA [90%CI] = .19, .36; CFI = .91; GFI = .93). The fit indices were slightly low. Modification indices suggested allowing error variance to correlate between Item 32 and 27. This improved the model fit (χ^2 = 3.26, *df* = 1, *p* = .07; RMSEA [90%CI] = .19, .36; CFI = .99; GFI = .99). The third measurement model for ID had good fit indices (χ^2 = 9.02, *df* = 2, *p* = .01; RMSEA [90%CI] = .00, .26; CFI = .96; GFI = .96).

Next, the new measurement model for ETMC with the three new testlets was tested using LISREL. The new testlets NO, TN, and ID were set to the second-order construct ETMC. The initial model fit needed improvement (χ^2 = 296.40, *df* = 41, *p* < .001; RMSEA [90%CI] = .17, .21; CFI = .90; GFI = .76). The fit indices were too low. Modification indices suggested allowing error variance to correlate between Item 23 and Item 28, between Item 27 and Item 32, and between Item 35 and Item 37. These modifications were made as item content showed significant conceptual relationships. The modified model had an improved fit (χ^2 = 129.57, *df* = 38, *p* < .001; RMSEA [90%CI] = .09, .14; CFI = .97; GFI = .88). This model can be seen in Figure 4.

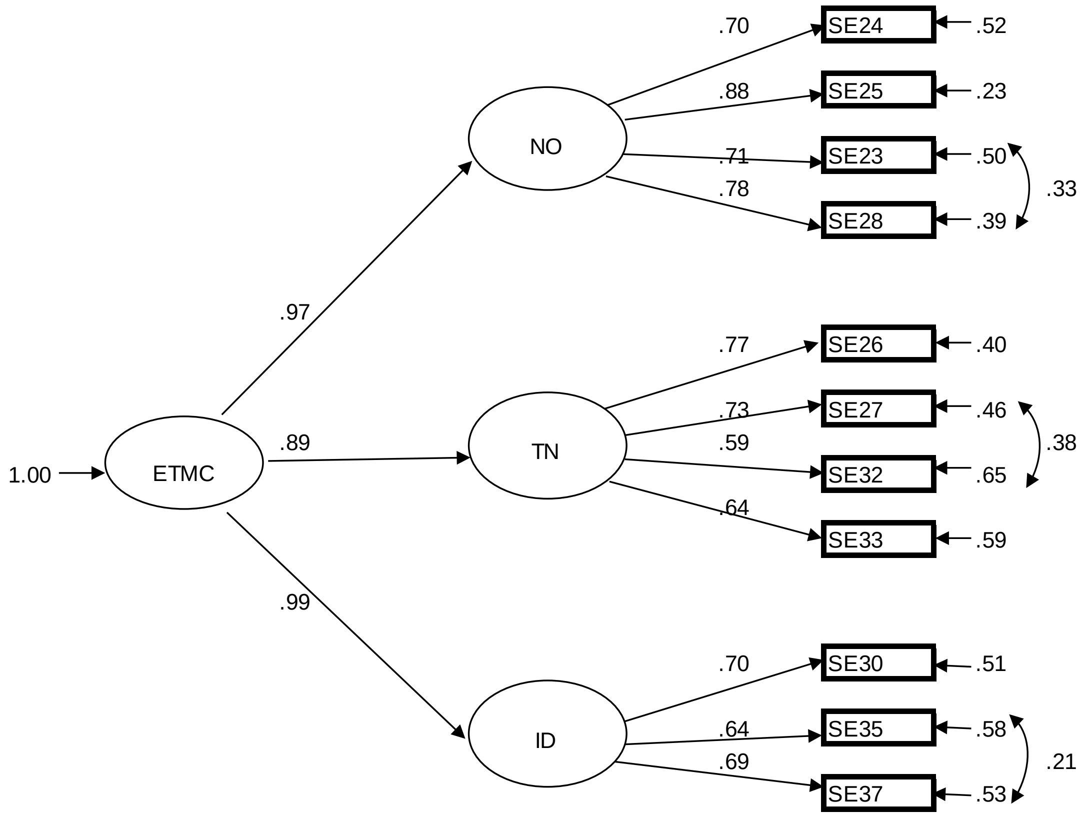

Figure 4. Structural model results of second-order construct ETMC with three testlets: *number and operations* (NO), *transforming numbers* (TN), and *interpreting data* (ID).

The measurement model was tested again for MTSE since Kindergarten teachers were removed from the analysis. The initial model fit needed improvement (χ^2 = 98.20, *df* = 14, *p* < .001; RMSEA [90%CI] = .15, .22; CFI = .92; GFI = .86). The fit indices were too low. Modification indices suggested allowing error variance to correlate between Item 1 and Item 2 and between Item 6 and Item 7. These modifications were made as item content showed significant conceptual relationships. The modified model had an improved fit (χ^2 = 36.43, *df* = 12, *p* < .0005; RMSEA [90%CI] = .06, .14; CFI = .98; GFI = .95). This model can be seen in Figure 5.

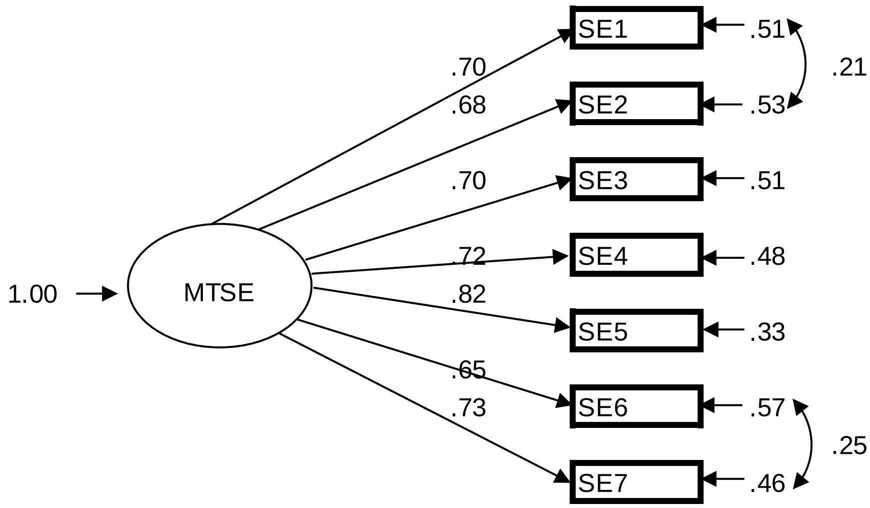

Figure 5. Structural model results of second-order construct MTSE without Kindergarten.

The final structural model (Model 3) was tested with kindergarten removed from the dataset. The initial model fit needed improvement as fit indices were too low (χ^2 = 448.19, *df* = 125, *p* < .001; RMSEA [90%CI] = .09, .12; CFI = .94; GFI = .81). LISREL did not display a path diagram as the model was non-admissible. Model 4 can be viewed in Figure 6. The fit indices can be seen in Table 13 compared against Model 1 and Model 2.

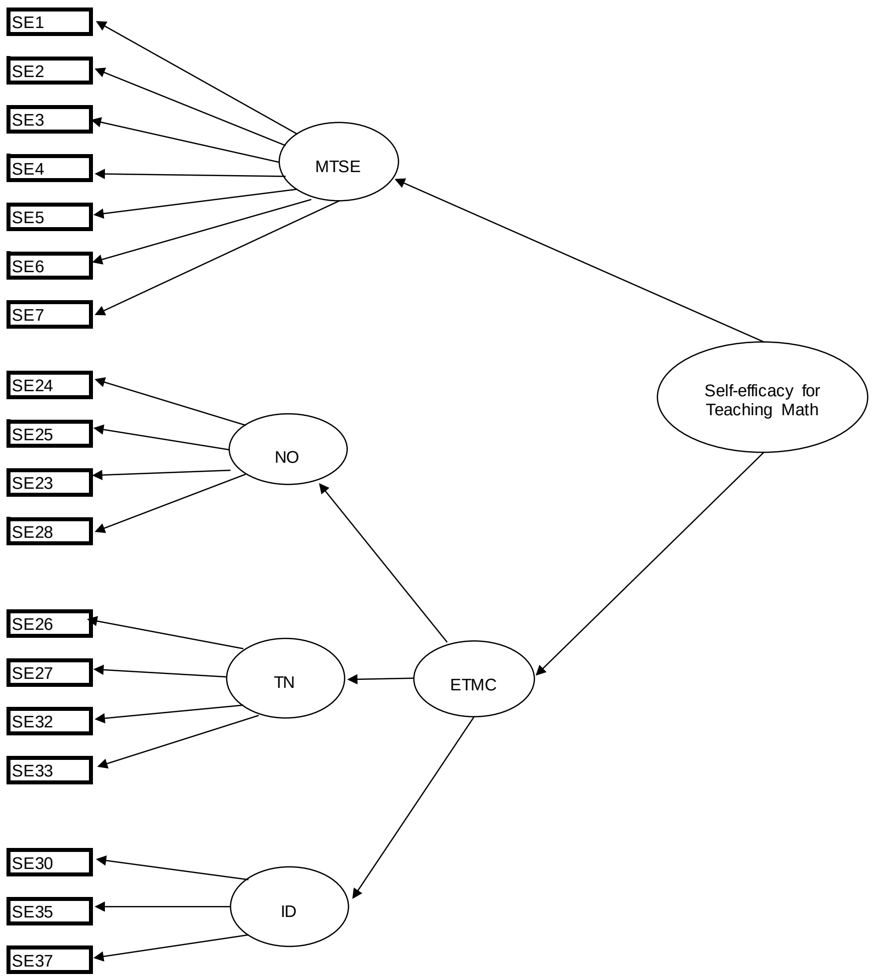

Figure 6. Structural model of Self-efficacy for Teaching Mathematics without Kindergarten.

Self-concept for Mathematics Content Knowledge

As mentioned earlier, Items 8 through 22 were not analyzed in the structural model for self-efficacy as the item content may actually be a better measurement of self-concept for mathematics content knowledge. These items fit the covariance matrix when left in the full structural model, but they needed to be further examined without Kindergarten teachers in the dataset. Upon examining the measurement model of these items without the Kindergarten data, the initial model did not fit the covariance matrix (χ^2 = 609.21, *df* = 90, *p* < .001; RMSEA [90%CI] = .16, .18; CFI = .91; GFI = .71). The modification indices suggested allowing error covariances to correlate between Item 9 and Item 10; Item 11 and Item 12; and Item 17 and Item 18. The modified model had an improved fit (χ^2 = 383.84, *df* = 87, *p* < .001; RMSEA [90%CI] = .12, .15; CFI = .95; GFI = .79). This model can be seen in Figure 6.

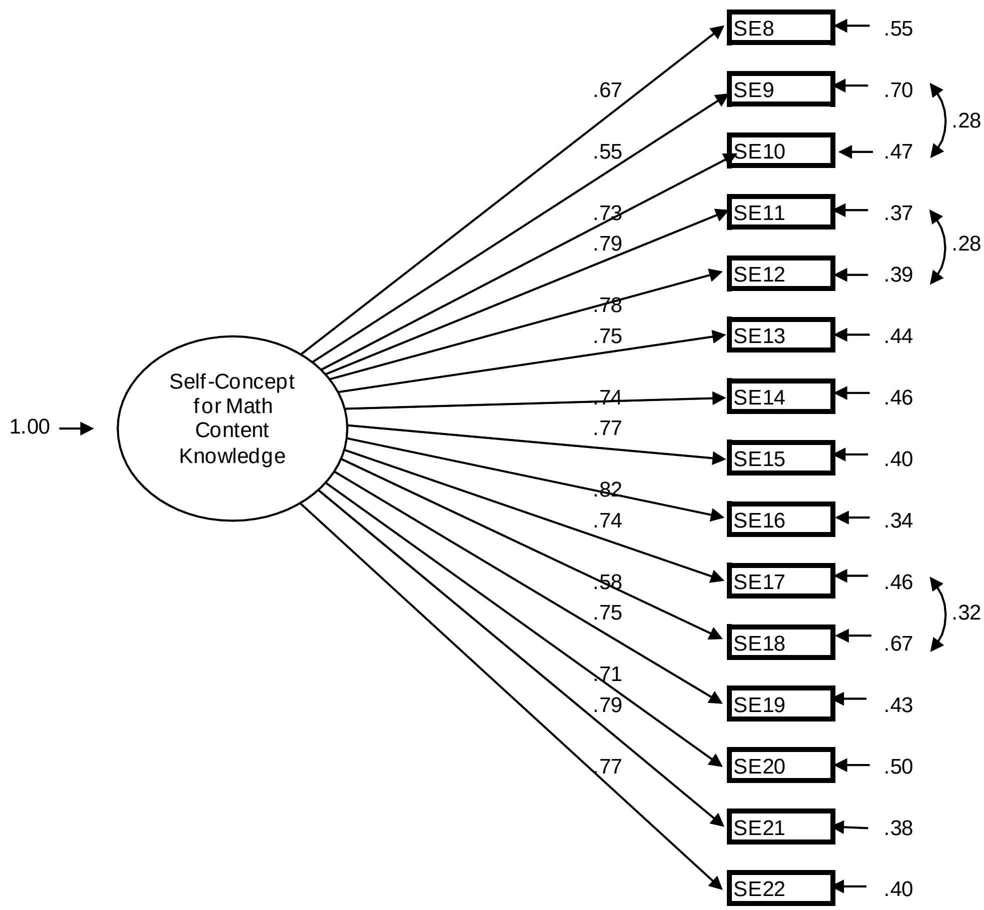

Figure 7. Measurement model results of Self-Concept for Mathematics Content Knowledge without Kindergarten.

Chapter Summary

This chapter explained in detail the results of statistical procedures and revisions made to the SETMI during this study. Reliability evidence was sound and concurrent validity provides promising evidence of the items in the SETMI as maintaining good measure of the universal concept of mathematics teaching self-efficacy. Reliability data were also promising. The EFA of Version One of the SETMI highlighted a more complex factor structure than was intended and called into question the validity of some of the items. The three construct model was kept intact, with sub-factors appearing as a result of the EFA.

The SETMI was then revised to both simplify the factor structure and to align mathematics content items more closely with the state standard course of study. Version One items were created as a result of consultation with mathematics content experts, but it appeared as though some items may have been too complex to accurately measure the constructs. Version Two items were evaluated by a focus group of practicing elementary teacher-leaders who thought the simplified items were concise and clear. The state standard course of study was consulted because elementary teachers in this state are licensed to teach in grades Kindergarten through six and it was important that the scope of content in those grades be represented.

Version Two of the SETMI was then evaluated by conducting a CFA. The three-construct model with a second-order construct of SETM was tested first. This model failed to fit the covariance matrix within reason and showed a large χ^2 and RMSEA that could not be meaningfully adjusted using the suggested modification indices. The CFI and GFI were too small to show evidence of adequate fit. In an effort to hone the original model, items were evaluated as to their alignment with Bandura's theory of self-efficacy. Upon close examination, EUMC items were deemed to be observable measures of mathematics content knowledge and not of self-efficacy for understanding mathematics content. These items were excluded from further analysis, although they may be good indicators of mathematics content knowledge that could lend itself to a structural model of self-efficacy.

The modified model was tested, with testlets being formed of items within the original ETMC construct. Items were divided into either *efficacy for teaching operations* (ETMO) or *efficacy for teaching concepts* (ETMC). The modified model consists only of three latent variables MTSE, ETMO, and ETMC. This model had a much improved fit and seemed promising.

In an effort to better understand the complexity of an elementary mathematics teacher's self-efficacy belief system, a MANOVA with grade as the independent variable and the three original constructs as dependent variables was conducted. There was a statistically significant different between groups for EUMC and ETMC, but not for MTSE. As noted previously, items for EUMC may not be measures of self-efficacy, therefore ETMC was again examined. An ANOVA was conducted with grade level as the independent variable and ETMC as the dependent variable. Kindergarten was statistically significantly different from grades 3 through 5 and was therefore removed from the data for further analysis. An EFA was then conducted on the new data set to determine if the removal of Kindergarten caused a changed in the factor structure. Three factors were revealed within ETMC and these became testlets that were further analyzed.

The measurement model for testlet NO was examined using LISREL. Model fit was acceptable, but allowing one modification caused the model to have improved fit. Testlet TN did not have acceptable model fit initially, but allowing one modification caused the model to fit the data. Testlet ID did have acceptable fit initially, but the model was still significant and one observed variable had a large error. This item (34) was deleted from the model, which caused ID to have perfect fit.

The structural model for ETMC was then tested using LISREL. The second-order factor ETMC was defined by testlets NO, TN, and ID, who were further defined by their related observed variables. This model did not fit the data without allowing for three modifications. These modifications were to allow error variances to correlate between items within the testlet. The fit indices for this model are much improved from Models 1 and 2, but still have statistical significance.

Finally, the mathematics content items (8 through 22) were analyzed separately as they may in fact be a better measure of self-concept for mathematics content knowledge. The initial model did not fit the covariance matrix, but upon allowing for logical correlations between items the model had improved fit. The inclusion of these items into a structural model for self-efficacy may be considered as they may contribute to pedagogical content knowledge.

CHAPTER 5: DISCUSSION

The theoretical basis for this study aligns with Bandura's (1977) ideas about self-efficacy from Social Cognitive Theory. Self-efficacy is one's beliefs about how their actions produce given future attainments (Bandura, 1977; Tschannen-Moran & Woolfolk Hoy, 2001). Currently there is one widely accepted measure of teacher self-efficacy, the Teacher's Sense of Efficacy Scale (TSES) (Tschannen Moran & Woolfolk Hoy, 2001) and one widely accepted instrument to measure the mathematics self-efficacy beliefs of teachers, the Mathematics Teaching Efficacy Beliefs Instrument (MTEBI) (Enochs et al., 2000).

The TSES has been tested repeatedly for validity using factor analysis. The factor structure has remained consistent over time. However, it has not been tested using confirmatory factor analysis, therefore the construct validity of the TSES has not been tested using structural methods. The MTEBI is used widely, and was determined to be valid by its creators, but its validity has been called into question in recent literature (Kieftenbeld, et al., 2010). The need for a mathematics-specific self-efficacy instrument to measure the beliefs of teachers is evident and was the guiding purpose for this study.

In response to the need for better measurement of teacher self-efficacy within a mathematics context, the Self-efficacy for Teaching Mathematics Instrument (SETMI) was created. The SETMI is comprised of three parts: Mathematics Teaching Self-efficacy

(MTSE), Efficacy for Understanding Mathematics Content (EUMC), and Efficacy for Teaching Mathematics Content (ETMC). Reliability for each part of the SETMI was found to be acceptable for both versions of the instrument. Additionally, concurrent validity of the SETMI was explored by correlating EUMC and ETMC to MTSE. Correlations were all statistically significant at the .01 level and therefore the SETMI was assumed to have acceptable concurrent validity.

To examine factor structure, the first EFA of Version One of the SETMI revealed a more complex factor structure than was anticipated. Revisions to Version One of the instrument were made to simplify the factor structure, clarify item content, and align mathematics content items to the state Standard Course of Study. Following other recommendations from the focus group, the SETMI was revised and administered to a second cohort of participants. A confirmatory factor analysis (CFA) was conducted on Version Two of the instrument to examine the construct validity.

The initial model did not fit the data, but a modification to the model excluding EUMC (Items 8 and 22) and creating testlets within the ETMC construct proved fruitful. Further analysis of the ETMC construct into smaller testlets proved to be a better model fit, along with the exclusion of Kindergarten teachers from the data. The iterations of the theoretical model for ETMC gave insight into the potential nature of an elementary mathematics teacher's self-efficacy beliefs.

Proposed Structural Model of Self-efficacy

Although the EUMC construct was removed in the Confirmatory Factor Analyses of Models 2 and 3, the items should not be completely abandoned. The EUMC construct items (8 through 22) were analyzed separately with Kindergarten removed to check the construct validity of those items. Although it was concluded that those items were not an accurate measure of self-efficacy for content knowledge, they may represent self-concept for mathematics content knowledge. Self-concept could influence self-efficacy as self-concept of content knowledge could be a component of pedagogical content knowledge. Pedagogical Content Knowledge (PCK) is a part of a mathematics teacher's belief system (Ernest, 1989; Peterson, Fennema, Carpenter, Loef, 1989), although it influences self-efficacy.

With the self-concept model considered separately from the self-efficacy model, and the relationship between the two acknowledged, it seems that the nature of self-efficacy beliefs may in fact be a structural model. The findings from the final CFA suggest that there is a bit more complexity to self-efficacy beliefs than may have first been realized by researchers. The impact of the specific components of mathematical content on the factor structure of ETMC suggests a close relationship to Ernest's (1989) ideas about a mathematics teacher's complex belief systems. In further examination of these items they are referenced as "self-concept for mathematics content knowledge." This proposed structural model of SETM can be found in Figure 7.

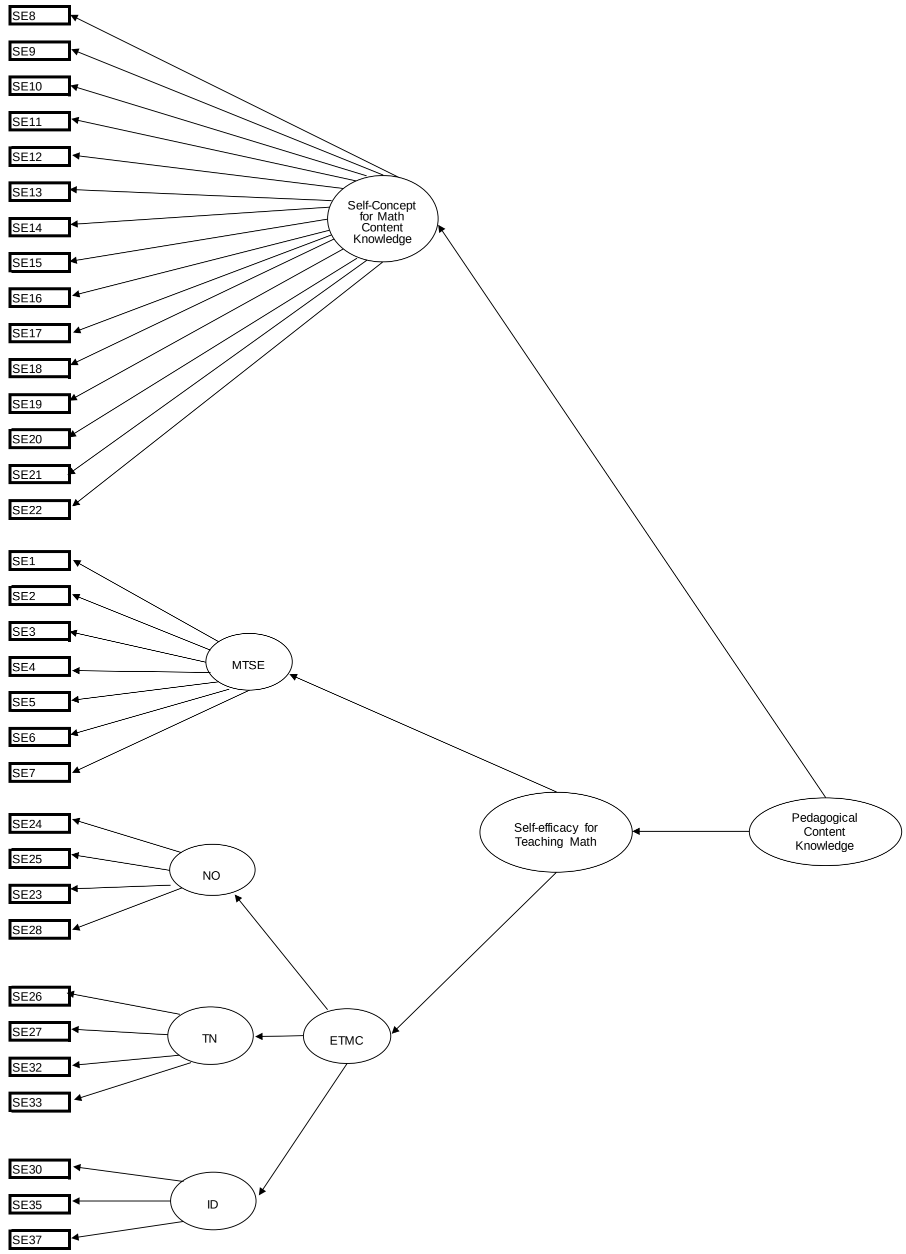

Figure 8. Proposed structural model of elementary mathematics teacher self-efficacy that includes self-concept for mathematics content knowledge and pedagogical content knowledge.

Ultimately, these results show evidence that the SETMI is a valid and reliable measure of two aspects of self-efficacy: *teaching mathematics* and *teaching mathematics* *content*. Modification of the initial theoretical model gave great insight into the potential for self-efficacy to be much more complex than was initially thought. Research on self-efficacy measurement has attempted to capture the construct through the formation of instruments aligned with tightly formed constructs thought to be elements of self-efficacy. These earlier instruments clung tightly to Bandura's ideas that self-efficacy is a component of *efficacy expectations.* Bandura (1997) also stated that the effect of self-concept on self-efficacy is weak, but he did not quantify the strength of the potential relationship. Additionally, using mathematics as a context for examining self-efficacy creates a need to examine self-concept for mathematics content knowledge since this is integral to the ability to teach mathematics successfully. Bandura (1997) stated that personal attributes may or may not be relevant to their efficacy for completing a task or producing an outcome, but in the case of teaching mathematics it seems logical that personal attributes could contribute heavily to both one's decision to teach mathematics and to one's self-efficacy for doing so.

The relationship between teacher's beliefs and practices is also important to reiterate. Numerous studies have noted a relationship between teacher beliefs and practices (Beswick, 2012; Ernest, 1989; Stipek et al., 2001). Therefore, a teacher's beliefs are important to understanding their practices. Since there is also a direct relationship between teacher practices and student learning (Darling-Hammond & Youngs, 2002), the belief system of a teacher is incredibly important to understand. If student achievement in mathematics is to improve, the nature of a mathematics teacher's complex belief system must be understood.

The proposed model in this study provides evidence that perhaps a structural model of self-efficacy is more accurate and that measurement of self-efficacy has to include more components than just self-efficacy for teaching mathematics content and self-efficacy for teaching practices in mathematics. There may be an interaction with pedagogical content knowledge that has not been explained in this study. These and other factors may contribute to self-efficacy in a way not previously described in the literature.

Implications for Practitioners

There are several implications for practitioners regarding the use of a specific measure of self-efficacy for elementary mathematics. First, measuring self-efficacy in such a specific context can allow for specific feedback. In elementary education, the focus of this study, it is assumed that the teacher is an expert in all academic areas. While this might be the case for some elementary teachers, many feel as though they lack confidence and skill in mathematics (McGee, Wang, & Polly, in press; Piel & Green, 1993; Polly, McGee, Wang et al., 2010). If teachers could be administered a measure that would provide specific feedback about their levels of self-efficacy in teaching mathematics, they might gain a greater understanding of their own belief systems. Additionally, specific professional development could be provided to increase self-efficacy beliefs.

Providing specific professional development to teachers, based on areas of mathematics teaching where they might be weaker is both beneficial to the school administration and the teacher. Attending only professional development focused on areas of mathematics teaching where a teacher needs support is a much more efficient use of time and resources. Mathematics content itself is extremely complex. Building a teacher's skill set in one area may help improve their skill in another, thus increasing their self-efficacy.

Increasing an elementary teachers' self-efficacy beliefs should have a direct impact on their teaching practice, which is another implication for this study. If student achievement in mathematics is an area of focus, a teacher's practices in mathematics teaching should also be examined. There is a known relationship between student achievement in mathematics and teacher practice. There is also a known relationship between a teacher's beliefs and their practice. Providing specific support to the teacherfor areas of mathematics teaching where they might be less efficacious could increase the achievement of their students. Measuring teacher self-efficacy and student performance could aid in the understanding of that relationship.

Implications for Researchers

Researchers in both psychological and educational contexts have a vested interest in the measurement of self-efficacy. Although much effort was given to the development of these instruments in the past, work on creating modern instruments to measure the concept has stalled. Even widely accepted instruments that have been used to measure self-efficacy should be re-examined as measurement literature has evolved to introduce new ideas related to establishing validity evidence. Most instruments that are in use today within the literature are at least 10 years old, which indicates a need to reexamine both validity and reliability of previously used scales. For example, Confirmatory Factor Analysis has seldom been used to examine the construct validity of existing instruments, even those widely used in the literature. A failure to thoroughly examine instrumentation before using it to make conclusions in research can cause the measurement of underlying constructs to be inaccurate.

Context specific measurement of self-efficacy in particular has been difficult to obtain for mathematics (Swackhammer, 2010). Mathematics content and teaching practices are both broad fields that are difficult to assess at a granular level. When existing measures of self-efficacy have been modified to be mathematics specific, it has been difficult to maintain the existing factor structure of the instrument. Even existing mathematics teaching self-efficacy instruments may not provide the depth or breadth necessary to adequately measure this concept.

Beyond the need to find ways to measure self-efficacy at the utmost levels of specificity, there lies a need to understand the concept more deeply. Previous models of self-efficacy measurement were concise and tightly bound to developed constructs. It is perhaps more logical that self-efficacy is complex structural model with parameters that have yet to be defined. This study provides a glimpse into what that might look like, which will allow for greater understanding of the concept and more specific ideas about measurement.

Finally, this study brings to light a potential relationship between self-efficacy for teaching mathematics and pedagogical content knowledge (PCK). Although there is some research examining beliefs and PCK, educational psychology research has not crossed over into content knowledge research to bridge this gap. This leaves for an exciting field of inquiry for future research on self-efficacy.

Limitations of the Study

As with any study, there are limitations to be considered. First, the data used in this study, although representing a relatively large population of elementary teachers, comes from participants of a MSP grant who applied to participate in a grant focused on increasing teacher knowledge about mathematics practices. Additionally, the SETMI was intended only to be used to measure the self-efficacy beliefs of inservice teachers. Therefore, the common practice of using self-efficacy instruments with preservice teachers should be considered with caution. The triadic nature of the theoretical model presented here may not be appropriate for pre-service teachers as their experience is limited. Finally, the generalizability of the instrument to other populations of teachers should not be assumed. Content items in the SETMI are specific to elementary teachers and may not be good items to measure self-efficacy of middle grades or secondary teachers.

Need for Future Research

This study highlights several needs for future research. First, the conceptual model that is the underpinning for most currently used instruments may not have remained stable over time, as the pressures of teaching have shifted to a focus on accountability of teachers. Additionally, the job expectations of a teacher have also changed to encompass many social and developmental responsibilities for their students. This evolution in the profession may affect general teacher self-efficacy.

Most previously published instruments use the RAND items as a foundation and there is some question as to how well that aligns to current thoughts about self-efficacy. These instruments usually claim to measure between two and three constructs related to personal and teaching efficacy. However, Henson (2001) and others have made claims that these constructs may tie too closely with Rotter and not Bandura. An examination of previously published instruments is inherent to ensuring that quality research continues to be published in this area.

With regard to context specific self-efficacy, more development work needs to be done to add useful instruments to the literature. The focus on context-specific self-efficacy allows for differentiation within populations of teachers, or within one teacher's experience. It is entirely possible that when examining specific contexts, a teacher's self-efficacy levels may differ. It is also possible that factors which may have not been previously thought to be related to self-efficacy could in fact have an impact on self-efficacy development and sustainability. Self-efficacy of a teacher contributes both to job satisfaction and to student performance, which impacts school resources and professional development opportunities.

Another distinction not well drawn in the literature is the difference between the self-efficacy levels of pre-service and in-service teachers. Too many studies use widely accepted measures of in-service teacher efficacy (MTEBI, TSES, STEBI) to measure pre-service teacher efficacy. Although several authors have warned against doing so, this mis-use of instrumentation for populations for which the instrument was not created causes much confusion in the literature. If the self-efficacy level of pre-service teachers is to be measured, it is vital to consider their varying levels of experience and create appropriate instruments to examine self-efficacy in meaningful ways.

More work is also needed to examine PCK within elementary mathematics teachers. As mentioned in Chapter 4, the self-concept model did not fit the covariance matrix without modifications. Other work could be done to improve the items in this construct so that self-concept could be included in the structural model of self-efficacy.

Finally, specific work on the instrument that is the focus of this study is invited. The concept of self-efficacy within teaching is much more complex than originally accounted for in the literature. More work should be done to expand the structural model for self-efficacy to include pedagogical content knowledge, mathematical content knowledge, years of experience, and grade levels taught to ensure that measurement of the concept is accurate and precise.

REFERENCES

American Educational Research Association [AERA]. (1999). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Aerni, P. (2008). *Teacher self-efficacy and beliefs for teaching mathematics in inclusive settings*. (Doctoral dissertation). Retrieved from Proquest. (Publication No. AAT 3353198).

Aiken, L. R. (1970). Attitudes toward mathematics. *Review of Educational Research, 40,* 551-596.

Armor, D. J., Conry-Oseguera, P., Cox, M. A., King, N., McDonnell, L. M., Pascal, A. H., … Thompson, V. M. (1976). *Analysis of the School Preferred Reading Program in Selected Los Angeles Minority Schools*. (Report No. R-2007-LAUSD). Santa Monica, CA: RAND.

Askew, M., Brown, M., Rhodes, V., Johnson, D., & William, D. (1997). *Effective teachers of numeracy: Final report*. London: King's College.

Ashcraft, M. H., & Moore, A. M. (2009). Mathematics anxiety and the affective drop in performance. *Journal of Psychoeducational Assessment, 27,* 197−205.

Ashton, P. T., Olejnik, S, Crocker, L. & McAuliffe, M. (1982). *Measurement problems in the study of teachers' sense of efficacy*. Paper presented at the annual meeting of the American Educational Research Association, New York.

Aston, P. T., & Webb, R. B. (1986). *Making a difference: Teacher's sense of efficacy and student achievement.* New York: Longman.

Baldwin, W. (2009). Information no one else knows: The value of self-report. In Stone, A. A., Turkkan, J. S., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (Eds*). The science of self-report: Implications for research and practice* (3-8). Mahwah, NJ: Taylor & Francis.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.

Bandura, A. (1986). *Social foundations of thought and action.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1994). *Self-efficacy*. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (Vol. 4, pp. 71-81). New York: Academic Press.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bandura, A. (2006). *Guide for Constructing Self-efficacy scales.* In. Pajares, F. & Urdan, T. (Eds.). *Self-efficacy beliefs of adolescents*. Greenwich, CT: Information Age Publishing,

Berman, P., McLaughlin, M., Bass, G., Pauly, E., & Zellman, G. (1977). *Federal Programs Supporting Educational Change: Vol. 7. Factors Affecting Implementation and Continuation*. Santa Monica, CA: RAND.

Beswick, K. (2012). Teachers' beliefs about school mathematics and mathematicians' mathematics and their relationship to practice. *Educational Studies in Mathematics, 79*, 127-147. doi: 10.1007/s10649-011-9333-2

Betz, N. E. & Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *Journal of Vocational Behavior, 23*, 329-345.

Betz, N. E. & Hackett, G. (1986). Applications of self-efficacy theory to understanding career choice behavior. *Journal of Social and Clinical Psychology, 4,* 279-289.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15,* 1-40.

Brualdi,A. (1999). Traditional and Modern Concepts of Validity.(Research Report No. EDO-TM-99-10). Retrieved from http://eric.ed.gov/PDFS/ED435714.pdf

Bursal, M., & Paznokas, L. (2006). Mathematics anxiety and preservice elementary teachers confidence to teach mathematics and science. *School Science and Mathematics, 106*, 173–179.

Carpenter, T.P., Fennema, E., Franke, M.L., Levi, L., & Empson, S.B. (2000). *Cognitively Guided Instruction: A Research-Based Professional Development Program for Elementary School Mathematics*. Report No. 003. Wisconsin Center for Educational Research: Madison, WI.

Comrey. A. L., & Lee, H. B. (1992). *A first course in factor analysis (2nd Edition).* Hillsdale, NJ: Lawrence Erlbaum Associates.

Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Darling-Hammond, L. & Youngs, P. (2002). Defining “highly qualified teachers”: What does “scientifically-based research” actually tell us? *Educational Researcher, 31*, 13-25.

Dawes, J. G. (2002). Five point vs eleven point scales: Does it make a difference to data characteristics? Australian Journal of Market research, 10, 1-47.

Dawes, J. G. (2008). Do data characteristics change according to the number of scale points used? International Journal of Market Research, 50, 61-77.

DeCoster, J. (1998). Overview of Factor Analysis. Retrieved September 5, 2011 from http://www.stat-help.com/notes.html

Denzine, G. M., Cooney, J. B., & McKenzie, R. (2005). Confirmatory factor analysis of the Teacher Efficacy Scale for prospective teachers. British Journal of Educational Psychology, 75(4), 689-708. Doi:10.1348/000709905X37253

Dornyei, Z. (2009). Motivation and motivating in the foreign language classroom. *The Modern Language Journal, 78*, 273-284.

Ebel, R. L. (1980). Survey research in education: The need and the value. *Peabody Journal of Education, 57,* 126-134.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.* Ft Worth, TX: Dryden Press.

Enochs, L. G., Smith, P. L., & Huinker, D. (2000). Establishing Factorial Validity of the Mathematics Teaching Efficacy Beliefs Instrument. *School Science and Mathematics, 100,* 194-202.

Ernest, P. (1989). The impact of beliefs on the teaching of mathematics. In P. Ernest (Ed.), *Mathematics Teaching: the State of the Art* (pp. 249–54). London: Falmer.

Fan, X., & Sivo, S. A. (2005). Sensitivity of fit indices to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling, 12,* 343-367.

Fennema, L., Carpenter, T., Franke, M., Levi, M., Jacobs, V., & Empson, S. (1996). A longitudinal study of learning to use children’s thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27,* 403-434.

Gay, L. R., Mills, G. E., & Airasian, P. (2009). *Educational Research: Competencies for analysis and applications*. Upper Saddle River, NJ: Pearson Education, Inc.

Gibson, S. & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76,* 569-582.

Green, M., Piel, J. A., & Flowers, C. (2008). Reversing education majors' arithmetic misconceptions with short-term instruction using manipulatives. *The Journal of Educational Research, 101,* 234-242.

Guskey, T. (1981). Measurement of responsibility teachers assume for academic successes and failures in the classroom. *Journal of Teacher Education, 32,* 44-51.

Haines, P., McGrath, P., & Pirot, M. (1980). Expectations and persistence: An experimental comparison of Bandura and Rotter. *Social Behavior and Personality, 8*, 193-202.

Harrell, A. V. (1985). Validation of self-report: The research record. In Rouse, B. A., Kozel, N. J., & Richards, L. G. (Eds.). *Self-report methods of estimating drug-use: Meeting current challenges of validity* (pp. 12-21). Rockville, MD: National Institute on Drug Abuse.

Heneman, H. G., Kimball, S., & Milanowski, A. (2007). *The Teacher Sense of Efficacy Scale: Validation Evidence and Behavioral Prediction*. WCER Working Paper No. 2006-7.

Henson, R. K. (2001, January). *Teacher self-efficacy: Substantive implications and measurement dilemmas*. Paper presented at the annual meeting of the Educational Research Exchange, Texas A&M University, College Station, TX.

Hill, H.C., Schilling, S.G., & Ball, D.L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal, 105*, 11-30.

Hill, H.C., Rowan, B., & Ball, D.L. (2005) Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*, 371-406.

Hoffman, B. (2010). "I think I can, but I'm afraid to try": The role of self-efficacy beliefs and mathematics anxiety in mathematics problem solving efficiency. *Learning and Individual Differences, 20*, 276-283.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6, 1-55.

Huang, S. C. & Chang, S. F. (1998). Self-efficacy in learners of English as a second language: Four examples. *Journal of Intensive English Studies, 12,* 23-40.

Huck, S. W. (2008). Reading Statistics and Research. Boston, MA: Allyn and Bacon.

Johnson, B., & Christensen, L. (2004). *Educational Research: Quantitative, Qualitative, and Mixed Approaches (second edition).* Boston, MA: Allyn and Bacon.

Kahle, D. K. B. (2008). *How elementary school teachers' mathematical self-efficacy and mathematics teaching self-efficacy relate to conceptually and procedurally oriented teaching practices* (Doctoral dissertation). Retrieved from Proquest. (1555882761)

Kranzler, J. H., & Pajares, F. (1997). An exploratory factor analysis of the mathematics self-efficacy scale revised. *Measurement and Evaluation in Counseling and Development*, *29*, 215-228.

Kieftenbeld, V., Natesan, P., & Eddy, C. (2010). An item response theory analysis of the mathematics teaching efficacy beliefs instrument. Journal of Psychoeducational Assessment, doi:10.1177/0734282910391062

Klassen, R. M., Bong, M., Usher, E. L., Chong, W. H., Huan, V. S., Wong, I. Y. R. & Georgiou, T. (2009). Exploring the validity of a teachers' self-efficacy scale in five countries. *Contemporary Educational Psychology*, *34*, 67-76, doi: 10.1016/j.cedpsych.2008.08.001.

MacCallum, R. C., Roznowski, M. & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*, 490 – 504.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320-341.

McGee, J., Wang, C., & Polly, D. (in press). Guiding Teachers in the use of a Standards-Based Mathematics Curriculum: Teacher Perceptions and Subsequent Instructional Practices after an Intensive Professional Development Program. Article to Appear in *School Science and Mathematics*.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741-749.

Mokros, J. (2003). Learning to reason numerically: The impact of Investigations. In S. Senk & D. Thompson (Eds.), *Standards-Based School Mathematics Curricula:*

*What Are They? What Do Students Learn? (*pp. 109-131). Mahwah, NJ: Lawrence Erlbaum Associates.

Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research, 62,* 229-258.

Nolan, D. (2009). *A study of the relationship of teachers' self-efficacy and the impact of professional learning community as an organizational structure.* (Doctoral dissertation, Louisiana State University). Retrieved from http://etd.lsu.edu/docs/available/etd-06102009-104808/unrestricted/Nolan.pdf

Ozer, D. J. (1985). Correlation and the coefficient of determination. *Psychological Bulletin*, *97*, 302-315.

Pajares, F. & Miller, M. D. (1994). The role of self-efficacy and self-concept beliefs in mathematical problem-solving: A path analysis. *Journal of Educational Psychology*, *86*, 193-203.

Pajares, F., & Miller, M. D. (1995). Mathematics self-efficacy and mathematics performances: The need for specificity of assessment. *Journal of Counseling Psychology, 42,* 190-198.

Pajares, F. (1997). Current directions in self-efficacy research. In M. Maehr & P. R. Pintrich (Eds.). *Advances in motivation and achievement*.Volume 10, (pp. 1-49). Greenwich, CT: JAI Press.

Pajares, R. (2002). *Overview of social cognitive theory and of self-efficacy*. Retrieved from http://www.des.emory.edu/mfp/eff.html

Peterson, P. L., Fennema, E., Carpenter, T. P., & Loef, M. (1989). Teachers' pedagogical content beliefs in mathematics. *Cognition and Instruction*, *6*, 1-40.

Phelps, C. M. (2010). Factors that pre-service elementary teachers perceive as affecting their motivational profiles in mathematics. *Educational Studies in Mathematics, 75*, 293-309.

Piel, J. A. & Green, M. (1993). Attitudes toward five subject areas: A comparison of elementary education and noneducation majors. *North Carolina Journal of Teacher Education, 6*, 28-37.

Polly, D., McGee, J., Wang, C., Lambert, R.G., Pugalee, D.K. & Johnson, S. (2010). *The association between teacher beliefs and practices and student learning in mathematics.* Manuscript submitted for publication.

Riggs, I., & Enochs, L. (1990). Towards the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education, 74*, 625-637.

Rose, J.S., & Medway, F. J. (1981). Measurement of teacher's beliefs in their control over student outcome. *Journal of Educational Research, 74*, 185-190.

Ross, J., & Bruce, C. (2007). Professional Development Effects on Teacher Efficacy: Results of Randomized Field Trial. *Journal of Educational Research, 101*, 50-60.

Rotter, J. B. (1954). *Social learning and clinical psychology.* New York: Prentice-Hall.

Scholz, U., Gutierrez-Dona, B., Sud, S., & Schwarzer, R. (2002). Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *European Journal of Psychological Assessment, 18*, 242-251.

Schunk, D. H. (1994). Self regulation of self-efficacy and attributions in academic settings. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (pp. 75-99). Hillsdale, NJ: Erlbaum.

Schumacker, R. E. & Lomax, R. G. (2010). *A Beginner's Guide to Structural Equation Modeling*. New York, NY: Taylor & Francis.

Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist, 26*, 207-231.

Schwarzer, R. & Born, A. (1997). Optimistic self-beliefs: Assessment of general perceived self-efficacy in 13 countries. *World Psychology, 3,* 177-190.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences (3rd edition).* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Swackhamer, L. E. (2010). *Measuring mathematics specific teacher efficacy: Can a global instrument produce valid results?* Paper presented at the 2010 American Educational Research Association Annual Meeting, Denver, CO.

Swan, M. (2006). Designing and using research instruments to describe the beliefs and practices of mathematics teachers. *Research in Education, 75*, 58-70.

Tabachnick, B. G. & Fidell, L. S. (2007). *Using Multivariate Statistics*. Boston, MA: Allyn and Bacon.

TERC (2009). *Investigations in Number, Data, and Space*. Pearson.

Thompson, B. & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement, 56*, 197-208.

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783-805.

Tschannen-Moran, M., & Woolfolk Hoy, A. (2005). The differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and Teacher Education, 23*, 944-956.

Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, 68, 202-248.

United States Department of Education (2008). *The final report of the National Mathematics Advisory Panel*. Washington, DC: Author.

Vick, R. M., & Packard, B. W. (2008). Academic success strategy use among community-active urban Hispanic adolescents. *Hispanic Journal of Behavioral Sciences, 30*, 463-481.

Wainer, H., Bradlow, E. T., Wang, X. (2007). *Teslet response theory and its applications*. New York, NY: Cambridge University Press.

Wang, C., & Pape, S. (2007). A probe into three Chinese boys' self-efficacy beliefs learning English as a second language. *Journal of Research in Childhood Education, 21*, 364-377.

Wenglinsky, H. (1999). *Does it compute? The relationship between educational technology and student achievement in mathematics.* Educational Testing Service Policy Information Center. Retrieved from www.mff.org/pubs/ME161.pdf.

Wold, S., Esdensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *2*, 37-52.

Wolters, C. A., & Daugherty, S. G. (2007). Goal structures and teachers' sense of efficacy: Their relation and association to teaching experience and academic level. *Journal of Educational Psychology, 99*, 181-193.

Woolfolk, A. E., & Hoy, W. K. (1990). Prospective teachers' sense of efficacy and beliefs about control. *Journal of Educational Psychology, 82*, 81-91.

Zeldin, A. L. & Pajares, F. (2000). Against the Odds: Self-Efficacy beliefs of women in mathematical, scientific, and technological careers. *American Educational Research Journal, 27*, 215-246. doi: 10.3102/000283120370012

APPENDIX A: SELF-EFFICACY FOR TEACHING MATHEMATICS INSTRUMENT (VERSION ONE)

Name: ____

<u>Directions</u>: Please circle the number that matches your response.

None at All	Very Little		Strong Degree		Quite A Bit		A Great Deal	
1	2	3	4	5	6	7	8	9

1.	To what extent can you motivate students who show low interest in mathematics?	1	2	3	4	5	6	7	8	9
2.	To what extent can you help your students' value learning mathematics?	1	2	3	4	5	6	7	8	9
3.	To what extent can you craft relevant questions for your students related to mathematics?	1	2	3	4	5	6	7	8	9
4.	To what extent can you get your students to believe they can do well in mathematics?	1	2	3	4	5	6	7	8	9
5.	To what extent can you use a variety of assessment strategies in mathematics?	1	2	3	4	5	6	7	8	9
6.	To what extent can you provide an alternative explanation or example in mathematics when students are confused?	1	2	3	4	5	6	7	8	9
7.	How well can you implement alternative teaching strategies for mathematics in your classroom?	1	2	3	4	5	6	7	8	9
8.	How much time do you feel you have to plan to teach mathematics?	1	2	3	4	5	6	7	8	9
9.	To what extent do you effectively teach mathematics?	1	2	3	4	5	6	7	8	9

Not Confident	Barely Confident	Somewhat Confident	Confident	Very Confident
1	2	3	4	5

How confident do you feel in your understanding of:

10.	Integers, rational, and irrational numbers.	1	2	3	4	5
11.	The equivalence of fractions, decimals, and percents.	1	2	3	4	5
12.	Arithmetic operations on decimals and fractions.	1	2	3	4	5
13.	Solving one to two step arithmetic word problems.	1	2	3	4	5
14.	Inverse relationships between multiplication and division.	1	2	3	4	5
15.	Coordinate planes.	1	2	3	4	5
16.	Interpreting bar and line graphs.	1	2	3	4	5
17.	Use of compasses, rulers, and protractors.	1	2	3	4	5
18.	Square and cubic units.	1	2	3	4	5
19.	Size, quantity, and capacity.	1	2	3	4	5
20.	Use of websites to promote mathematical understanding.	1	2	3	4	5

21.	Use of estimation as a problem-solving strategy.	1	2	3	4	5
22.	Identifying, describing, and creating patterns.	1	2	3	4	5
23.	Solving one and two-step equations.	1	2	3	4	5
24.	Different representations to describe a proportional relationship.	1	2	3	4	5

How confident do you feel in your ability <u>teach students</u> to:

25.	Understand integers, rational, and irrational numbers.	1	2	3	4	5
26.	Describe equivalence of fractions, decimals, and percents.	1	2	3	4	5
27.	Perform arithmetic operations on decimals and fractions.	1	2	3	4	5
28.	Solve one to two step arithmetic word problems.	1	2	3	4	5
29.	Understand inverse relationships between multiplication and division.	1	2	3	4	5
30.	Locate points on a coordinate plane.	1	2	3	4	5
31.	Interpret bar and line graphs.	1	2	3	4	5
32.	Use compasses, rulers, and protractors.	1	2	3	4	5
33.	Understand square and cubic units.	1	2	3	4	5
34.	Measure size, quantity, and capacity.	1	2	3	4	5
35.	Use of websites to promote mathematical understanding.	1	2	3	4	5
36.	Use estimation as a problem-solving strategy.	1	2	3	4	5
37.	Identify, describe, and create patterns.	1	2	3	4	5
38.	Solve one and two-step equations.	1	2	3	4	5
39.	Use different representations to describe a proportional relationship.	1	2	3	4	5

APPENDIX B: SELF-EFFICACY FOR TEACHING MATHEMATICS INSTRUMENT (VERSION TWO)

Name: ____________________________________

Grades You Have Taught (circle): K 1 2 3 4 5

None at All	Very Little	Strong Degree	Quite a Bit	A Great Deal
1	2	3	4	5

Directions: Please circle the number that matches your response.
Your individual responses will not be shared.

1.	To what extent can you motivate students who show low interest in mathematics?	1	2	3	4	5
2.	To what extent can you help your students’ value learning mathematics?	1	2	3	4	5
3.	To what extent can you craft relevant questions for your students related to mathematics?	1	2	3	4	5
4.	To what extent can you get your students to believe they can do well in mathematics?	1	2	3	4	5
5.	To what extent can you use a variety of assessment strategies in mathematics?	1	2	3	4	5
6.	To what extent can you provide an alternative explanation or example in mathematics when students are confused?	1	2	3	4	5
7.	How well can you implement alternative teaching strategies for mathematics in your classroom?	1	2	3	4	5

How well do you understand:

8.	Characteristics of Numbers (i.e. whole numbers, rational/ irrational numbers).	1	2	3	4	5
9.	Strategies for composing and decomposing numbers by manipulating place value in addition and subtraction.	1	2	3	4	5
10.	Strategies for composing and decomposing numbers by manipulating place value in multiplication and division.	1	2	3	4	5
11.	Conversion of a fraction to a decimal and vice versa.	1	2	3	4	5
12.	Comparing equivalence of fractions and decimals	1	2	3	4	5
13.	Inverse relationships between operations (i.e. +, - and *, ÷)	1	2	3	4	5
14.	Coordinate planes.	1	2	3	4	5
15.	Collecting, plotting and interpreting data (on any type of graph)	1	2	3	4	5
16.	Measurement of area and perimeter	1	2	3	4	5
17.	Conversions between units in the same system (i.e. grams → kilograms, inches → yards).	1	2	3	4	5
18.	Conversions between units in a different system (i.e. kilograms → pounds, inches → centimeters).	1	2	3	4	5
19.	Measuring the length of objects.	1	2	3	4	5
20.	Mathematical patterns	1	2	3	4	5
21.	Variables in an algebraic equation.	1	2	3	4	5
22.	Probability of outcomes	1	2	3	4	5

How well can you teach students to:

23.	Describe characteristics of Numbers (i.e. whole numbers, rational/ irrational numbers).	1	2	3	4	5
24.	Perform strategies for composing and decomposing numbers by manipulating place value in addition and subtraction.	1	2	3	4	5
25.	Perform strategies for composing and decomposing numbers by manipulating place value in multiplication and division.	1	2	3	4	5
26.	Convert a fraction to a decimal and vice versa.	1	2	3	4	5
27.	Compare equivalence of fractions and decimals	1	2	3	4	5
28.	Interpret inverse relationships between operations (i.e. +, - and *, ÷)	1	2	3	4	5
29.	Manipulate coordinate planes.	1	2	3	4	5
30.	Collect, plot and interpret data (on any type of graph)	1	2	3	4	5
31.	Measure area and perimeter	1	2	3	4	5
32.	Convert between units in the same system (i.e. grams → kilograms, inches → yards).	1	2	3	4	5
33.	Convert between units in a different system (i.e. kilograms → pounds, inches → centimeters).	1	2	3	4	5
34.	Measure the length of objects.	1	2	3	4	5
35.	Discover and create mathematical patterns	1	2	3	4	5
36.	Interpret variables in an algebraic equation.	1	2	3	4	5
37.	Interpret probability of outcomes	1	2	3	4	5