UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

VALIDATION OF A MEASURE OF TEACHERS' EFFICACY AND OUTCOME EXPECTATIONS IN THE CONTENT DOMAINS OF READING AND MATHEMATICS

A Dissertation

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

By

LESLIE KAY CURDA
Norman, Oklahoma

1997

UMI Number: 9733890

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

VALIDATION OF A MEASURE OF TEACHERS' EFFICACY AND OUTCOME EXPECTATIONS IN THE CONTENT DOMAINS OF READING AND MATHEMATICS

A Dissertation APPROVED FOR THE
DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

BY

ACKNOWLEDGMENTS

There are many people to which I would like to express my gratitude for continually encouraging my professional and personal growth, challenging me at every turn, and making my years at the University of Oklahoma exciting and productive.

I would like to thank Ray Miller, the chair of my committee and my advisor from the beginning. He provided continual support for the development of my professional interests in both educational psychology and educational technology. He smiled and nodded each time I wanted to change my research interests or my dissertation topic. He patiently accepted each new deadline I set and reset for myself without ever questioning whether I could do it. He introduced me to the theory of self-efficacy, challenged my thinking, provided critical feedback, and was the ideal mentor for me.

I would like to thank my committee members to whom I am forever grateful for making my experiences enjoyable. Teresa DeBacker, who began as my mentor for my first college teaching experience and continued throughout my four years to provide support and guidance in developing my teaching and research interests. Paul Kleine, who provided continuous support and encouragement through his storytelling and who always understood my thinking because, deep down, he thinks like a graduate student and a teacher. Everett Smith, who challenged me in every course, fine-tuned my knowledge of measurement and statistics, and introduced me to Rasch methodology. Bonnie Konopak, who in her quiet way always let me know that she was proud of me and that I could accomplish whatever I decided to tackle.

I would like to thank other faculty members in the EIPT program, Tim Ragan, Pat Smith, Barbara Greene, and Jay Smith, who have contributed to my professional development through their courses, guidance, and support.

I would also like to thank my family. My parents, Will and Brenda, who have always encouraged my intellectual growth, are always truly interested in my current projects, and are never short on research project ideas for the future. My older brother and his family, Allen, Ann, Caroline, Katherine, and Elizabeth, who are one of the primary reasons I ventured to the University of Oklahoma for my graduate studies and made my first year considerably less lonely than I thought it would be. My older sister, Lee, and her husband, Joe, who have provided encouragement and housing for many conferences throughout my four years. My younger brother and his family, David, Mendy, Megan, and Kaylee, who trekked out west with me and always had a bed for me in Dallas when I needed to get away.

Most importantly, I would like to thank my husband, Steve, who is my best friend and most fervent supporter. He patiently understood whenever I became difficult to live with when deadlines were approaching. He played both mother and father to our daughter in order for me to have entire days to work without interruption. He always knows what to say and, more importantly, what not to say, when I need encouragement or a little boost to my self-esteem. Without him in my life there would be no sun in my sky.

Finally, I owe the preservation of my sanity throughout the last year to my daughter, Riley. She understood when mommy needed to work by herself in the study but also knew when I needed a break. Her quiet knocks on the closed door of the study and quiet calls to me through the door periodically reminded me that it

was time to enjoy the finer things in life -- watching her learn to crawl and walk, sharing in each of her new discoveries (even if it was a crack in the sidewalk), walking Shakespeare around the block, playing in the park, being a family, sharing a Coke Icee, and dancing, among other things.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This study provided evidence of the reliability and validity of inferences from a newly developed instrument measuring teachers' efficacy and outcome expectations in the content domains of reading and math. The instrument developed is more closely aligned with Bandura's construct of expectancy than previously developed instruments and other commonly used measures of teacher efficacy and outcome expectations in the research literature. Results of the content validity studies provided solid evidence that (a) the conceptual and operational definitions of teacher efficacy and outcome expectations were aligned with Bandura's expectancy theory, (b) the items could be identified as measuring four distinct constructs, and (c) the items represented a range of difficulty.

Rasch analyses identified a few items measuring each construct that were misfitting and subsequently eliminated. Final results provided evidence for construct validity through the final item statistics for each construct and the hierarchy of item difficulty. The Rasch results also indicated areas in which the instruments may need improvement, including creation of items assessing the higher end of the efficacy and outcome continuums in both reading and math and further examination of the functioning of the response scale categories. Acceptable measures of the internal consistency of responses to each set of items measuring a construct and estimates of the precision of the instruments in consistently measuring person ability were also found.

Confirmatory factor analysis procedures failed to produce similar findings as the Rasch procedures. This was an interesting, but not entirely unexpected result. The lack of congruence between the results of these two procedures,

however, did not call into question the construct validity, but rather provided further evidence for the need to carefully consider the analytic techniques used when conducting validation studies.

As hypothesized, both efficacy and outcome measures were correlated with measures of effort within their domain. The pattern of correlations also offered evidence for the need to distinguish between efficacy and outcome expectations and to measure them independently. Other criterion measures, such as teacher planning and teacher engagement, were not correlated with measures of efficacy and outcome expectations as hypothesized. More well-defined variables and use of multiple indicators of planning and engagement should be developed to further investigate these hypothesized relationships.

CHAPTER I

Introduction

Background of the Project

Bandura (1977) discussed the role of cognition in the acquisition and regulation of behavior and motivation. He introduced the two-dimensional construct of expectancy to explain and predict behavioral and motivational changes produced from different treatments or situations. His theory has been applied to a variety of research settings in education including teacher thinking and behaviors in the classroom. However, research in these contexts as well as others has often been plagued with measurement problems attributable to inappropriate operationalization of the constructs of efficacy expectations and outcome expectations.

Teacher efficacy expectations and teacher outcome expectations are possibly two of the most important social-psychological factors influencing teacher behaviors and student outcomes in the classroom. Empirical research in this field has begun to provide evidence that measures of teacher efficacy expectations and teacher outcome expectations are predictive of (a) teacher behaviors that may serve to enhance or hinder student engagement and learning in classrooms, (b) measures of student efficacy and goals in learning situations, and (c) measures of student cognitive engagement and achievement (Anderson, Greene, & Loewen, 1988; Armor et al., 1976; Ashton, 1985; Gibson & Dembo, 1984; Midgley, Feldlaufer, & Eccles, 1989; Roeser, Arbreton, & Anderman, 1993). However, as such evidence for the importance and role of teacher efficacy expectations and teacher outcome expectations in determining teacher and student behaviors builds, concern for the validity and reliability of the inferences made from instruments used in such research has also begun to surface. As is the case with any research, when the instruments used to measure the constructs under investigation come into question, so do the findings of studies using the instruments.

In this project I investigated the validity and reliability of inferences from a newly developed instrument measuring teacher efficacy expectations and teacher outcome expectations in the content domains of reading and math. I developed the instrument after careful review of Bandura's (1977) concept and theory of efficacy and outcome expectations and the measurement problems often associated with investigating these constructs. Next, I critically examined present definitions and measures of teacher efficacy expectations and teacher outcome expectations and conducted a review of empirical evidence of their associations with teacher behavior and student outcomes. Finally, I developed an instrument intending to measure teacher efficacy expectations and teacher outcome expectations in the content domains of reading and math.

In this dissertation I will describe the process of instrument development and validation. This process is one that is never complete as results of validation procedures often lead to further development and refinement of the instrument. The following steps describe the procedures I followed to complete this project and that I will report in this dissertation:

1. I completed a review of theory and research related to teacher efficacy and outcome expectations.

2. I developed an instrument to measure the constructs of teacher efficacy expectations and teacher outcome expectations in the content domains of reading and math.

3. I conducted content validity procedures including (a) assessment of the adequacy of the conceptual definitions as they relate to the proposed use of the instrument and assessment of the correspondence between the conceptual definitions and the instrument items by experts in the field of motivation in education, (b) analysis of the extent to which the items truly reflect the conceptual categories and the domains of interest by experts in the field of teaching and motivation in education, and (c) evaluation of instrument items for adequacy of the sampling from all possible instructional activities and for adequacy of sampling from a range of easy to difficult tasks by experts in the field of teaching. I analyzed the results after each of these procedures and made any necessary changes in the instrument prior to the next step of the content validation.

4. I completed a pilot administration of the instrument and analyzed the data through computation of means, standard deviations, ranges, and Cronbach alpha reliability coefficients to identify any initial administration problems or item weaknesses prior to final administration using a sample from the target population.

5. I administered the final version of the instrument to a random sample of 1500 elementary school teachers.

6. I conducted an analysis of the content, construct, and criterion-related validity; and reliability using (a) Rasch rating scale analysis procedures, (b) confirmatory factor analysis procedures, and (c) correlational analyses.

Significance of the Project

In this project I developed and examined the validity and reliability of inferences made from an instrument measuring teacher efficacy expectations and teacher outcome expectations in the content domains of reading and mathematics. My intention was to remedy early empirical research flaws in the measurement of teacher efficacy expectations and teacher outcome expectations through the construction of a measure more clearly in alignment with Bandura's theory. In addition, I developed the teacher efficacy expectations and teacher outcome expectations instrument within a specific context to meet the needs and purposes of continued research.

As with many states, there is mandatory testing of students from grades two through twelve in the state of Tennessee. Presently, I have access to the database which provides an assessment of teachers' past and present achievements through measures of their students' gain scores in reading and mathematics. While longitudinal analysis of this data is ongoing and findings are interesting, no additional data has been or is being collected to account for many of the findings. For example, some teachers consistently have high student gains each year in reading or math while others do not (even when controlling for school setting, size, ethnic composition, and a variety of other variables). Current theory and research on teacher efficacy suggests that it is quite possible that variations in some combination of teachers' efficacy and outcome expectations, and teachers' effort, persistence, and choice of instructional tasks could account for the reported variance in teachers' student gain scores in reading and math. However, before we can confidently draw conclusions from an investigation testing these ideas we need to have confidence that the inferences we would like to make from measures of teacher efficacy expectations and outcome expectations for the domains of interest are valid and reliable. The development of such a measure of teacher efficacy expectations and teacher outcome expectations in reading and math is the goal of this study.

Future research utilizing this instrument includes investigation and identification of variables that influence teacher efficacy expectations and teacher outcome expectations in the content domains of reading and math, and examination of the influence of teacher efficacy expectations and teacher outcome expectations on teacher effort, persistence, and choice or preference for instructional tasks; and student motivation and achievement.

CHAPTER II

Current Literature

Measures of teacher efficacy expectations and outcome expectations to predict teacher motivation and behavior, and student motivation and achievement, have been used as early as 1976. Most of this research cites Bandura's (1977) theory of expectancy as the framework for measuring teacher efficacy expectations and teacher outcome expectations and linking these constructs to a variety of other variables proposed by his theory. Although empirical research has successfully found relationships among teacher efficacy expectations and teacher outcome expectations, and teacher and student behaviors, recent literature has criticized the instruments used to measure teacher efficacy expectations and teacher outcome expectations. The areas of research reviewed below are: (a) Bandura's theory of efficacy and outcome expectations; (b) Bandura's theory applied to teaching, teachers, and students; (c) measurement issues related to Bandura's expectancy construct; and (d) current definitions and measures of teacher efficacy expectations and teacher outcome expectations.

Bandura's Theory of Efficacy and Outcome Expectations

Bandura (1986) described outcome expectations and efficacy expectations as instrumental in accounting for behavioral and motivational changes in the individual. A depiction of his theory is in Figure 1. Outcome expectations are beliefs that behaviors will or will not lead to desirable or valued outcomes. Such expectations indicate how certain one is that the desired consequence will result from successful performance of a task or implementation of a strategy. Efficacy expectations are beliefs that one does or does not possess the required skills to bring about the performance and are often referred to as self-efficacy in the literature. Bandura (1986) defined self-efficacy as an individual's judgment of his or her capability to establish and carry out behaviors required to achieve a specified type of performance. Self-efficacy is an individual's judgment of what he or she can do with the skills he or she possesses rather than a simple judgment of the skills he or she possesses.

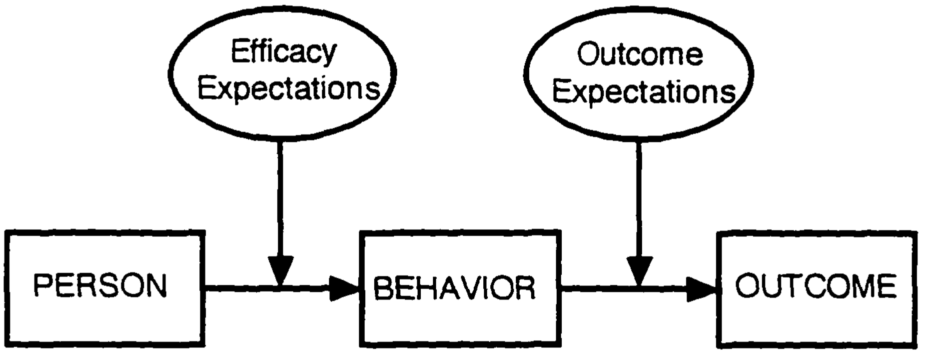

Figure 1. Graphical Depiction of Bandura's Expectancy Theory

Bandura (1986) stated both efficacy and outcome expectancies are needed to best predict behaviors and motivation. Behaviors are best predicted by considering both types of expectancy determinants especially under circumstances where both efficacy and outcome expectancies vary. Bandura differentiated between these two expectancies because individuals can believe that a particular action will produce a desired outcome (which would increase their motivation to perform), but they may not act on this belief because they do not possess confidence in their ability to actually execute the necessary behaviors (thus decreasing motivation to perform). He predicted that persons high on both variables would respond in assured and assertive ways while persons low on both variables would give up readily when results were not immediate.

Under normal circumstances, when individuals see outcomes as contingent on their ability to perform specific behaviors, they rely more heavily on efficacy judgments when deciding which course of action to pursue. For example, an individual will judge his or her certainty of making a free-throw (outcome expectation) as contingent upon his or her ability (efficacy expectation) to release the ball (behavior) appropriately. In such a case, Bandura (1986) notes that it is impossible to sever the expected outcome from the performance judgment upon which it is conditional. For activities such as these, where expected outcomes are highly dependent on efficacy judgments, knowledge of expected outcomes may not add much to the prediction of behavior. Bandura believes this makes knowledge of individuals' efficacy expectations better predictors of behaviors than knowledge of outcome expectations. For this reason, Bandura has elaborated more on the construct of efficacy expectations than on outcome expectations.

However, it may also be the case that certain specified outcomes are not believed to be inextricably linked to the adequacy of behaviors performed. For example, a teacher may not always believe that a student's achievement in math is a direct result of his or her ability to teach math. The teacher may, instead, believe that a student's effort or innate ability is the major determinant of math achievement. In such a case, expected outcomes are believed to be independent or loosely linked to the specified performance. Here, separate judgments of efficacy and outcome expectations can be assessed because the structural arrangement of teaching may often result in the same performance producing variable outcomes. Bandura, however, does not elaborate on how knowledge of expected outcomes in these situations will add to the prediction of behavior.

<u>Dimensions and Sources of Efficacy Expectations</u>

Bandura (1977; 1986) suggested several dimensions along which efficacy expectations can vary and identified factors influencing the development of efficacy. Efficacy expectations can vary in generality, magnitude, and strength. Generality refers to the extent to which efficacy is consistent across domains or is situation-specific. For example, is a teacher who is confident in his or her ability to teach reading to second graders just as confident in his or her ability to teach math to the same group of students or to teach reading to eighth graders, or is such confidence domain specific (reading only) and/or situation specific (second graders only)? The extent to which behaviors, for which efficacy is judged, are broadly or narrowly defined is the generality of efficacy expectations.

The magnitude (or level) of efficacy expectations refers to the extent to which efficacy is limited to or inclusive of simple tasks, moderately challenging tasks, and/or the most difficult tasks or situations. In other words, when tasks are ordered in level of difficulty, is an individual only confident in successfully performing the simpler tasks or does his or her confidence extend to moderately challenging tasks and even the most difficult tasks (Bandura, 1977)? The strength of efficacy expectations refers to the degree of certainty about successful performance (Zimmerman, 1996) and influences the ease or difficulty with which they can be modified (Denham & Michael, 1981). Weak efficacy expectations may be easy to reduce through disconfirming experiences while strong efficacy expectations for success may produce increased efforts and persistence despite disconfirming experiences (Bandura, 1977).

Bandura (1977; 1986) also noted four major sources of efficacy information for an individual. Efficacy expectations are developed through (a) past experiences of success and failure, (b) vicarious experiences of similar others and their successes and failures, (c) verbal persuasion of credible others, and (d) physiological indices. This efficacy information only becomes useful through cognitive appraisal. Bandura (1986) suggests that factors including personal, social, situational, and temporal circumstances under which events occur will affect how experiences are cognitively appraised and serve to influence efficacy judgments. This cognitive appraisal includes two functions. The first function concerns the types of information a person attends to and uses as indicators of efficacy. The second function concerns the heuristics an individual uses for weighting and integrating efficacy information from different sources to make efficacy judgments.

The most influential source of efficacy is the individual's own performance attainments. Cognitive appraisal of these performance attainments includes (a) attributing successes or failures to external or internal factors, (b) assessing the difficulty of the task and the amount of effort exerted, (c) assessing the amount of external aid one receives and the circumstances under which he or she performed, and (d) the temporal pattern of successes and failures. Successes often serve to increase efficacy while failures, especially those attributed to ability and uncontrollable factors, often serve to decrease efficacy expectations. Mastery of a difficult or newly performed task is more likely to raise efficacy than success at an easy task or a task successfully performed repeatedly in the past. Individuals often view effort as inversely related to ability; thus success with minimal effort may serve to increase efficacy, especially if the task was challenging. If an individual fails but believes that little or no effort was exerted then the experience may do little to adjust efficacy judgments. An individual's self-monitoring of performances may also influence how experience contributes to efficacy judgments. Persons that selectively attend to negative performances are likely to underestimate their efficacy, while persons who note and remember their successes are likely to have higher efficacy judgments.

The successes and failures of others influence the development of self-efficacy in a similar fashion as personal performance attainments, especially when the person being observed is judged by the individual to be very similar in skills and abilities. Generally, successes by similar others raise efficacy and failures by similar others lower efficacy. Cognitive appraisal of the similarity of a model may focus on a model's past performances or a model's attributes that are perceived to be predictive of the performance about which efficacy is being judged. Appraisal of the similarity of a model's past performances are most useful when old and new activities are identical and situational demands are invariant. Appraisal of the similarity of a model's attributes may also focus on age, gender, SES, race, etc. due to preconceptions based on cultural stereotyping and overgeneralization. Several models with a variety of different attributes who demonstrate success on challenging tasks will serve to increase efficacy more than the same performance by a single model. Models may also enhance efficacy through teaching effective strategies for tasks. Models who fail through the use of an inappropriate strategy may also raise efficacy if an individual perceives that he or she knows a better strategy.

Efficacy expectations can also be developed through attempts of others to verbally convince an individual that he or she does possess the abilities needed to perform the identified task. Cognitive appraisal of the persuader is the key influence on efficacy judgments. A persuader who is credible or who presumably possesses the knowledge to evaluate another's competence will be most likely to influence efficacy judgments. Evaluative feedback from others who are skilled themselves in the activity or who use some objective predictors of performance attainments are likely to affect efficacy judgments.

Involuntary behaviors such as sweating or a racing heartbeat before or during a performance may be interpreted by an individual as signals that he or she is incapable of performing successfully. Cognitive appraisal may include indicators such as the source of arousal, the level of arousal, the circumstances under which arousal is elicited, and past experiences of how the arousal affected performance. The source of arousal is often established through social labeling processes. How one interprets sweating, for example, will have different impacts on efficacy. Interpretation of sweating as due to an uncomfortable temperature instead of fear may serve to keep efficacy judgments constant rather than lower them. Some people may view arousal as a performance enhancing indicator while others may believe that arousal adversely affects their performance. These views are often developed through past experiences of how arousal affected performances.

Finally, how an individual integrates these four sources of efficacy information in forming his or her efficacy judgments is a factor to consider. One may rely more heavily on past performance attainments in one situation and verbal persuasion in another situation while also considering all other sources. The sources themselves might also contradict one another and one must decide which source(s) to emphasize when making final estimates of efficacy. Bandura's emphasis on efficacy expectations and their sources has resulted in more research on this dimension of the expectancy construct.

Effects of Efficacy Expectations on Behavior

Research on efficacy expectations has concluded that people who maintain different levels of self-efficacy behave differently (Ashton, 1984; 1985; Ashton, Webb, & Doda, 1983; Bandura, 1986). Efficacy expectations can affect people's choice of or preference for activities, the amount of effort they will exert in a given activity, and their level of persistence when faced with difficulties (Bandura, 1977). Bandura (1986) stated that those who perceive themselves as highly efficacious set challenges that occupy their interest and have high involvement in activities of their liking. Those who possess high self-efficacy tend to put forth continuing efforts when they notice their performances fall short of their goals. They also tend to approach potentially threatening tasks nonanxiously. Their high self-efficacy tends to motivate behavior that produces accomplishment. In contrast, those who possess low self-efficacy tend to shy away from difficult tasks, lack effort, and give up easily when faced with difficult tasks. They tend to dwell on their personal deficiencies and suffer from much anxiety and stress (Bandura, 1986).

Outcome Expectations

Bandura and those utilizing his theory as a framework for their research often give little attention to the role of outcome expectations in predicting the behaviors mentioned in the previous section. This seems to be a gross oversight when considering Bandura's definition of outcome expectations. An individual is unlikely to choose to put forth effort or persist at a task if he or she does not believe that the outcome will follow from successful execution of the specified task or behavior. Circumstances in which there is a true or perceived bias or chance system operating may result in the belief that behaviors other than effort or persistence determine outcomes. Also, systems in which outcomes are dependent on multiple sources of influence and not just an individual's behaviors may result in the belief that such behaviors are only weakly linked to outcomes. In such cases, behaviors of effort, persistence, and task engagement, are likely to be highly related to outcome expectations rather than efficacy expectations.

One of the reasons that researchers may choose to ignore the role of outcome expectations in predicting these behaviors is due to the difficulty of clearly distinguishing them from efficacy expectations for purposes of assessment and manipulation (Maddux, Norton, & Stoltenberg, 1986). These difficulties often result in researchers developing measures that confound efficacy expectations and outcome expectations so that it becomes unclear which measure is the better predictor of behaviors of interest (Manning & Wright, 1983). High correlations have often led researchers to drop outcome expectations as a variable in predicting behavior and focus on the utility of efficacy expectations, instead of attempting to assess the variables independently by operationalizing and measuring them in a manner more consistent with Bandura's conceptual distinction.

Bandura's Theory Applied to Teaching, Teachers, and Students

A variety of qualitative and quantitative research has been conducted on factors influencing the efficacy of teachers. The rationales for most of these studies were related to the hypothesized link between teacher efficacy and teacher motivation and behavior, and teacher efficacy and student motivation and achievement. The brief review of studies below represents the sum of the research evidence to support Bandura's theory and the influences of teacher efficacy expectations and teacher outcome expectations on teacher motivation and behavior and student motivation, cognitive engagement and achievement. In his model of reciprocal determinism Bandura (1977) posited a bi-directional relationship between efficacy and behavior and achievement. So, it may be that student achievement is also actually influencing teachers' efficacy and behaviors and student motivation and cognitive engagement. The relationships among these variables are most certainly complex.

The efficacy and outcome expectations teachers hold are thought to (a) directly influence their engagement in specific behaviors related to instructional practices and relations with students, (b) directly or indirectly (through students' interpretations of teachers' behaviors and engagement in those instructional activities) influence student motivation and goals for learning, and (c) directly or indirectly (through student motivation) influence student achievement. Teacher efficacy and outcome expectations are thought to directly affect preservice and practicing teachers' decisions about employing specific instructional strategies. Their expectancies of their ability to successfully employ a strategy and their students' ability to gain from the strategy may limit the number and types of instructional strategies they will consider in a particular situation or for a particular group of students. These choices of instructional practices are then proposed to influence student motivation, cognitive engagement and achievement. Motivated teachers are generally thought to have motivated students due to their ability to engage students in activities that are meaningful. Following this reasoning, motivated students who are engaged in meaningful activities are likely to have increased achievement.

Ashton (1985) proposed that "personal efficacy" and "teaching efficacy" together form a critical construct in explaining teacher motivation. From Bandura's theory, she expected that these variables would influence teachers' choice of learning activities, the amount of effort they expend in teaching, and the degree of persistence they maintain when confronted with difficulties. Through interviews of high and low efficacy teachers based on their responses to the two Rand items (discussed in a later section), Ashton (1984) found that teachers with high efficacy (a) feel their work is important and meaningful, (b) possess a stronger sense of responsibility to see that students learn, (c) expect their students to progress, (d) plan for student learning by setting goals and identifying strategies to achieve them, (e) feel good about themselves, their students, and their profession, and (f) are confident they can influence student learning. Teachers with high efficacy are more likely to persist in efforts to increase student achievement according to specified goals and engage in a wider variety of teaching strategies to reach all students. Teachers with low efficacy are less likely to engage in these behaviors that are likely to increase student learning.

Ashton (1984) found that low efficacy teachers (a) feel frustrated and discouraged about teaching, (b) expect their students to fail and misbehave, (c) feel students are responsible for their own learning, (d) do not consider goals and strategies when planning instruction, (e) are frustrated and have negative feelings about their work, and (f) do not involve students in the decision making process. These different teaching behaviors resulting from varying levels of efficacy have been found to influence teacher motivation (Ashton, 1984; 1985), student learning (Anderson, Greene, & Loewen, 1988; Ashton, 1985), and student efficacy and motivation (Midgley, Feldlaufer, and Eccles, 1989).

Teachers with low efficacy place responsibility for learning on the students. When students fail, these teachers look for explanations in nonschool factors, such as ability, motivation, or family background. In contrast, teachers with high efficacy take personal responsibility for students' learning. When students fail, these teachers examine their own performance and look for ways they might have been more helpful. The more responsible a teacher feels for ensuring student learning, the more likely he or she will be not to give up in the face of students' difficulty. This same teacher might also be less likely to convey to students that they are incapable or inadequate (Ashton, 1984).

Gibson and Dembo (1984) conducted a pilot study on four high efficacy and four low efficacy teachers to investigate the relationship between teacher efficacy and teacher behavior. They found that low efficacy teachers spent significantly more time in small group instruction than high efficacy teachers. Low efficacy teachers gave significantly more feedback in the form of criticism than high efficacy teachers. There was also a significant difference in lack of persistence with low-efficacy teachers being more likely to go on by giving the answer or asking another student after the initial student failed to answer a question correctly. While data are from a small sample of teachers, results indicate a possible relationship between teacher efficacy and teacher classroom behavior.

Roeser, Arbreton, and Anderman (1993) provided further evidence of the relationship between teacher efficacy and teacher behavior and student outcomes.

They studied the relationships among teachers' efficacy, pedagogical beliefs, instructional goals for students, and instructional practices, and their influences on students' goal perceptions of their class, goal orientation, self-efficacy, attitude toward school, and work avoidance and deep processing strategies. The teacher instrument measured teacher efficacy using the Rand items (Armor et al., 1976), items from Gibson and Dembo (1984), and some of their own items, but they did not distinguish between teacher efficacy expectations and teacher outcome expectations. The significant findings of Roeser et al. included that teacher efficacy was related to both the learning and performance beliefs and practices of teachers. Also, teacher efficacy beliefs were negatively correlated with student perceptions of performance goals in the classroom and their work avoidance strategies and positively correlated with student perceptions of their class as learning goal oriented.

Research by Armor et al. (1976) is the most cited study in the teacher efficacy literature and is credited as the first to indicate a relationship between teacher efficacy and student achievement. Interestingly, teacher efficacy was only a single variable among many that these researchers included to investigate prediction of reading achievement gains in minority students. Armor et al. conducted a study to identify aspects of school reading programs that were associated with substantial and consistent gains in standardized reading test scores among minority children. They surveyed teachers, principals, and reading coordinators in a sample of twenty schools using questionnaires to measure a variety of variables including teacher efficacy, classroom approaches to reading, parent contacts, principals' support, teachers' use of resources, and classroom atmosphere. They used regression analyses to investigate the influence of these variables on the reading gains of Black and Mexican students separately. The variables that were positively correlated with the reading gains of Black children were the extent of teacher training in the use of a variety of materials keyed to individual student needs, teachers' efficacy expectations, and the extent to which an orderly classroom was maintained. No specific factors were found to be correlated with reading gains for the sample of Mexican children.

Midgley, Feldlaufer, and Eccles (1989) studied teacher self-efficacy beliefs and their influences on student goal orientations and motivation. They followed 1329 students through grades six and seven to investigate whether changes in student goal orientations and motivation were related to changes in their teachers' efficacy beliefs. They found that teacher efficacy beliefs at the beginning of the year were predictive of students' perceptions of their ability and performance in math in the second semester of the school year. Teachers with low self-efficacy tended to have students who lowered their self-efficacy and perceptions of ability in math and increased their perceptions of the difficulty of math during the school year. They hypothesized that this may stem from teachers' self-efficacy beliefs being expressed through their behaviors in the classroom and students perceiving and adopting these same beliefs about themselves. They also found that students moving from high to low efficacy teachers during the transition from sixth to seventh grade had significantly lower expectations for success in math, lower perceptions of their performance in math, and higher perceptions of the difficulty of math at the end of seventh grade than those students who went from low to high efficacy teachers or those who experienced no change.

Ashton (1985) designed a study to investigate the relationships between teacher efficacy, teacher and student behavior, and student achievement. Correlational analyses found a significant relationship between teaching efficacy and student achievement on a math subtest when controlling for entering ability. Personal teaching efficacy expectations were significantly related to student achievement on the language subtest when controlling for entering ability. Similar findings were indicated by Moore and Esselman (1992). They found measures of teaching efficacy, using the scale by Gibson and Dembo (1984), were positively related to math achievement (a gain of approximately 3 months) in second and fifth grades.

Tracz and Gibson (1986) also used the Gibson and Dembo scale and other measures of teachers' use of time and student engagement and regressed them on students' scores on the California Test of Basic Skills. Correlational analyses found personal teaching efficacy significantly and positively correlated with reading achievement and teachers' use of whole group instruction, and it was negatively related to small group instruction. Teaching efficacy was positively correlated with language and math achievement. Teachers with high personal teaching efficacy tended to engage in whole group instruction while teachers with low personal teaching efficacy tended to engage in small group instruction. This supports Bandura's idea that people with varying levels of efficacy engage in different behaviors. Regression analyses indicated personal teaching efficacy as a significant predictor of reading achievement, teaching efficacy as a significant predictor of math achievement, and student engagement rate as a significant predictor of language achievement.

Collectively, these studies seem to provide evidence supporting the application of Bandura's theory to classroom teaching. However, there are several problems with the previous studies. Most of these problems are directly related to the measurement of teacher efficacy and outcome expectations. There were a wide variety of definitions of teacher efficacy and outcome expectations upon which instruments were based. Most of the instruments did not correspond with Bandura's theory. The researchers rarely provided evidence of validity and/or reliability of the inferences made from the instrument used. Often, researchers chose not to or were unable to separate the influence of efficacy expectations and outcome expectations in some of the measures.

Measurement of Efficacy and Outcome Expectations

Bandura (1986) emphasized that measures of efficacy expectations must be tailored to the domain of functioning being investigated and should include detailed assessment of the generality, magnitude, and strength dimensions. Studies comparing global and particularized measures of efficacy find that domain specific measures of efficacy surpass global measures in explanatory and predictive power (Bandura, 1977; Pajares, 1996a). Bandura stated that such methodology will permit a finer-grained analysis of the degree of congruence between efficacy expectations and individual behaviors. The greater the generality of the efficacy judgments assessed, the wider the range of behaviors efficacy measures will be able to predict. The stronger the efficacy expectations, the more likely persons are to persist at the behaviors associated with them and the more likely they are to perform them successfully (Bandura, 1977).

Bandura (1986) also noted that efficacy expectations are concerned with generative capabilities not component acts. Generative capabilities are described as ones "in which cognitive, social, and behavioral subskills must be organized into integrated courses of action to serve innumerable purposes" (Bandura, 1986, p. 391). Thus, teachers are not asked to assess the confidence in their ability to write on the chalkboard, arrange desks, greet children, etc. (component acts), but they are asked to assess their confidence in conducting whole group instruction or implementing a specified instructional strategy (generative capabilities).

Several researchers have noted the problems associated with building self-efficacy instruments and made suggestions for building more appropriate measures. Owen (1989) noted that because self-efficacy refers to relatively specific beliefs, anyone engaged in studying self-efficacy in a slightly different area must develop a new tool. Pajares (1996a) agreed and reasoned that when self-efficacy is assessed it should be consistent with and tailored to the specific outcomes or behaviors that are of primary interest. He showed that when precise judgments of capability match the specific outcomes measured then there is an increase in the accuracy of prediction. Pajares (1996b) noted that efficacy measures in educational research are often too global and general and lack specificity of measurement and consistency with the criterial task that optimizes the predictive power of self-efficacy beliefs. In addition, Zimmerman (1996) asserted that inaccurate self-judgments are common among a variety of subjects, including young children and novice and experienced subjects. In some cases misjudgment can be large, but this problem is not specific to efficacy measures.

In their review of currently available measures of self-efficacy in educational and psychological research, Vispoel and Chen (1990) noted that only 4.4% of the instruments measured all three components of self-efficacy (strength, generality, and magnitude) while the rest only measured strength. However, they also noted that when correlations between measures of strength and magnitude are reported, they are never lower than .90 indicating little need for having separate scales. In addition, they suggested that while measures of generality are important, it is difficult to come up with ways of assessing it and a comprehensive assessment would require a large scope validity study and may not be necessary for the intended purposes of some efficacy instruments.

Vispoel and Chen (1990) also found that outcome expectancy is seldom measured and suggested that inclusion of such a scale could provide useful information. Other weaknesses they noted were that 61.1% of the instruments reported no reliability data, 68.9% were based on small samples of 150 or less, and few provided any systematic evaluation of the validity of the instrument (e. g., test specifications to guide scale development, review of items by content experts, or prepiloting and revision of items). They warned that unless local norms and corresponding validity evidence are available for instruments caution should be used. The following review of educational research related to teacher efficacy and outcome expectations confirms that mismeasurement of these constructs is prevalent and should be addressed before further investigations are conducted on the influences of these expectancies on teacher behaviors and student outcomes.

Definitions and Measures of Teacher Efficacy Expectations and Teacher Outcome Expectations

Teacher educators, educational researchers, and others have investigated how the construct of expectancy can be used to account for teacher motivation and teachers' engagement in productive behaviors associated with a variety of student outcomes, such as motivation, cognitive engagement, and achievement. Many of these investigations have used Bandura's theory as the framework for their studies. A variety of definitions and measures have been developed. Some are more appropriate than others judging by their alignment with Bandura's concepts of efficacy and outcome expectancies. These definitions and measures are the focus of this section.

Teaching Efficacy, Personal Efficacy, and Personal Teaching Efficacy

Using Bandura's theory as a framework, Rand Corporation researchers defined teacher efficacy expectations and teacher outcome expectations in an effort to measure them (Ashton, 1985). They defined teacher outcome expectations as beliefs about whether or not their actions as teachers would affect or lead to desired student outcomes. Rand researchers used the extent to which teachers agreed or disagreed with the following statement to measure teacher outcome expectations:

> 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.

They defined teacher efficacy expectations as beliefs about whether or not they had the ability to carry out a specified action related to teaching. They used the extent to which teachers agreed or disagreed with the following statement to measure teacher efficacy expectations:

> If I really try hard, I can get through to even the most difficult or unmotivated students.

Their research discovered a link between teachers' responses on these two items and minority student achievement in reading (Armor et al., 1976). The definitions Rand researchers used seem to correspond well with Bandura's concepts of efficacy and outcome expectations, and their finding support Bandura's theory; but, there are several problems associated with their measures of teacher efficacy expectations and teacher outcome expectations.

Neither item specifically directed teachers to the behavior of teaching reading or the outcome of student achievement in reading. In the teacher outcome expectation item, the phrase "can't do much" leaves a lot of room for individual interpretation and does not indicate any specific behaviors by which the teachers can make their judgment, and "student's motivation and performance" is a very global and multidimensional outcome in the classroom. In the teacher efficacy expectation item, the phrase "if I really try hard," precludes teachers from being able to say that this is an easy rather than difficult task for them to perform. Additionally, being able to "get through" to the "difficult" students and the "unmotivated" students requires a great deal of individual interpretation of the abilities needed for success and likely requires different behaviors for these different students. It almost seems as if the teacher efficacy expectation item is actually measuring teacher outcome expectation more appropriately by asking if a behavior (trying hard) will lead to an expected outcome of "getting through" to a variety of students. The teacher outcome expectation item also seems to be measuring a teacher's belief about the link between home environment (a possible set of behaviors) and student motivation and performance (a set of possible outcomes). These problems indicate poor alignment of these items with Bandura's two-dimensional expectancy construct and call into question the validity of their findings linking measures of teacher efficacy expectations and teacher outcome expectations to student achievement.

Research by Ashton, Webb, and Doda (1983) used the same two Rand items to assess teacher efficacy to relate it to student achievement and teacher motivation and verify the findings of the Rand researchers. As an addition, they labeled and defined the two items as measuring two dimensions of teacher efficacy related to Bandura's expectancy construct. They also added and labeled a third dimension.

Teaching efficacy. Ashton, Webb, and Doda (1983) labeled their first dimension "teaching efficacy." They intended this construct to correspond to Bandura's outcome expectancy and the first Rand item above. They defined *teaching efficacy* as teachers' beliefs about the general relationship between teaching and learning. This definition is operationalized in the level of agreement to the statement, "These kids can't be motivated." This statement, however, is a poorly operationalized measure of Bandura's construct of outcome expectations. First, no behavior is specified that would possibly lead to the expected outcome of motivated kids. Therefore, the assumption must be made by the teacher that the statement is saying that there is no possible behavior that would lead to student motivation. As a teacher, I would think this statement would be difficult to agree with given that most teachers can think of at least one behavior that could possibly lead to student motivation. Agreement with this statement would likely indicate teacher hopelessness or helplessness rather than simply low outcome expectancy. Also, the Rand item measuring teaching efficacy seems to be asking about the relationship between the teacher (the person) and the outcome rather than the behavior and the outcome as was suggested by Bandura and depicted in Figure 1.

Ashton (1985) later redefined *teaching efficacy* as teachers' beliefs about their ability to affect student learning in spite of obstacles. This brings their concept of teaching efficacy further away from Bandura's outcome expectancy. Including obstacles into statements assessing teaching efficacy makes it difficult to determine the strength of the link between the behavior and outcome identified. While teachers' perceptions of obstacles to teaching are important, they are more appropriately assessed by asking teachers to what they attribute their low teaching efficacy (i. e., outcome expectations).

Personal efficacy. Ashton Webb, and Doda (1983) labeled their second dimension "personal efficacy" which they intended to correspond to Bandura's efficacy expectancy construct. They defined *personal efficacy* as teachers' personal sense of effectiveness as a teacher and operationalized it through the statement, "I can't motivate." Here, once again, there are several problems with the definition and operationalization of personal efficacy if the intention was to use Bandura's theory as a framework for the study of teacher efficacy. Efficacy expectations are not judgments of one's sense of effectiveness but judgments of what one can do with the skills one has. It is not a judgment of the effectiveness of a performance but a judgment of confidence that one does or does not possess the required skills to bring about the performance. The statement, "I can't motivate" is not assessing a teacher's confidence in the skills he or she possesses but actually seems to be assessing whether a teacher can produce a desired outcome (motivation). This statement, however, is not assessing an outcome expectation because it does not identify or assess any link between a performance and an outcome.

Personal teaching efficacy. Ashton, Webb, and Doda (1983) labeled their third dimension "personal teaching efficacy" and intended it to correspond to the second Rand item above. They defined *personal teaching efficacy* as an integration of personal efficacy and teaching efficacy as evidenced in the statement, "I can't motivate these kids." Unfortunately, this statement is not assessing anything different than the personal efficacy statement of "I can't motivate." By adding "these kids" to the statement, Ashton, Webb, and Doda are increasing the specificity of the statement and are possibly including a different level of magnitude (e. g., "these kids" might be more difficult to motivate than other kids), but they do not seem to be assessing an entirely different dimension as they have tried to indicate through the development of a third dimension.

Measurement and distinction of teaching efficacy, personal efficacy, and personal teaching efficacy. Research by Ashton, Webb, and Doda (1983) verified the need to distinguish between the two dimensions of teacher efficacy measured by the Rand items, *teaching efficacy* and *personal teaching efficacy*. In only one of their five samples did the two items significantly correlate with one another. Also, intensive interviews of teachers concluded that they clearly distinguished between teaching efficacy and personal efficacy. Regardless of these findings, the confusion over what is being measured and investigated is still a concern.

Ashton, Webb, and Doda (1983) acknowledged these three types of efficacy, used the two Rand items to measure two of them (teaching efficacy and personal teaching efficacy), and then discussed findings related to the third, personal efficacy. Additional confusion lies in calling all three "measures of efficacy" when their intent with at least one of these dimensions is to measure outcome expectations and not efficacy expectations. Ashton, Webb, and Doda's use of personal teaching efficacy as a measure of teacher efficacy seems even less appropriate as they continue to explain the importance of keeping the three dimensions distinct due to the likelihood that the appropriate teacher change strategy would depend on the origin of the sense of inefficacy.

Ashton and her colleagues believed teachers who possess low teaching efficacy often do not feel responsible for a factor over which they have no control so they tend to exert less effort in attempting to affect student behavior and achievement because they see the effort as futile. These teachers will also assume that their fellow teachers will not be able to affect those students. Because of this reasoning, teachers with low teaching efficacy are likely to experience little stress when they are unsuccessful in reaching their students and are able to maintain their self-esteem. These assumptions seem to fall more clearly under Bandura's conception of outcome expectations, but their operationalization of these assumptions through the statement "These kids can't be motivated" was poor.

Ashton and her colleagues also point out that other teachers may have low personal efficacy and feel they personally lack the ability to teach. Teachers in this situation experience high stress because they believe their ineffectiveness is due to a lack of personal ability rather than some uncontrollable factor within the student. This low personal efficacy will be especially affected if these teachers perceive other teachers in similar situations as being successful. Thus, a teacher who is confident in her ability to teach but doubtful of her students' ability to learn (high efficacy expectations and low outcome expectations) requires a different intervention than a teacher who is confident in her students' abilities to learn but doubtful of her ability to teach (low efficacy expectations and high outcome expectations) (Ashton, Webb, & Doda, 1983).

These explanations of the use of personal efficacy and teaching efficacy measures leads me to the conclusion that a measure of personal teaching efficacy is not helpful because of its inability to distinguish whether inefficacy is due to low teaching efficacy or low personal efficacy. Unfortunately, Ashton and her colleagues have not followed their own theory development related to teacher efficacy and outcome expectations. Instead, they continue to use measures which they propose to be assessing teaching efficacy and personal teaching efficacy. Regardless of the above problems and inconsistencies, several researchers have expanded on the research of the Rand Corporation, and Ashton and her colleagues, continuing to develop, validate, and use measures of "teacher efficacy" that concentrate on teaching efficacy and personal teaching efficacy rather than the more appropriate and theoretically sound teaching efficacy and personal efficacy.

Gibson and Dembo's (1984) Teacher Efficacy Scale

Gibson and Dembo (1984) developed and validated an instrument measuring teacher efficacy under the assumption that teacher efficacy consists of two separate dimensions. Their two dimensions were labeled teaching efficacy and personal teaching efficacy using Ashton, Webb, and Doda's (1983) terms. They defined teaching efficacy as the extent to which teachers believe the learning environment can be controlled, or the belief that any teacher's ability to bring about student learning is significantly limited by external factors, such as home environment and family background. Examples of their teaching efficacy items are, "If students are not disciplined at home, they aren't likely to accept any discipline" and "Even a teacher with good teaching abilities may not reach many students." These items, along with many of the others, are ambiguous and seem to be measuring different things. For example, the first item above seems to be asking teachers about the link between home discipline (possibly a behavior, but not a teacher behavior or performance) and the likelihood of students accepting discipline (a student outcome); the second item, however, seems to be assessing the link between a teacher (a person) and reaching students (an extremely ambiguous outcome). It is important to note, though, that these items do seem to match with Gibson and Dembo's own definition of teaching efficacy. The concern is that neither these items nor the definition of teaching efficacy is well aligned with Bandura's construct of outcome expectancy.

Gibson and Dembo (1984) defined personal teaching efficacy as teachers' convictions that they have the skills and abilities to bring about specific student outcomes. Examples of their personal teaching efficacy items are, "When a student does better than usual, many times it is because I exerted a little extra effort," and "If a student masters a new math concept quickly, this might be because I knew the necessary steps in teaching that concept." These items seem to correspond well with their definition of personal teaching efficacy. However, they are combining the two dimensions of efficacy expectations and outcome expectations into a single measure. In addition, it would seem that if a researcher is measuring two dimensions of teacher efficacy, one intending to measure Bandura's construct of outcome expectations and the other intending to measure a combination of Bandura's efficacy and outcome expectations, then these measures should correlate significantly with one another due to the overlap of assessment of outcome expectations in both. Gibson and Dembo's validation results indicate otherwise.

Gibson and Dembo (1984) developed The Teacher Efficacy Scale that consisted of 30 items on a 6 - point Likert scale. They provided clear evidence through factor analysis procedures for the two factors, teaching efficacy and personal teaching efficacy, using 16 of the original 30 items. The two factors were only moderately correlated ($r = -.19$). However, Gibson and Dembo are not measuring efficacy and outcome expectations according to Bandura's theory and definitions. Their definitions and items seem to correspond well with one another and their results indicate separate constructs, but these constructs lack congruence with Bandura's two-dimensional expectancy construct. In addition, Gibson and Dembo's instrument is a global measure of these constructs and lacks the specificity that Bandura advocated for measures of efficacy expectations.

Woolfolk and Hoy (1990) also used the two dimensions of teaching efficacy and personal teaching efficacy to examine the structure of preservice teacher efficacy. They used a modified version of the 16-item Teacher Efficacy Scale by Gibson and Dembo (1984). They used the 16 items and added the two Rand items and four of their own questions appropriate for preservice teachers to the scale. Factor analytic procedures resulted in a two factor solution consistent with the Gibson and Dembo (1984) findings. Others, such as Riggs and Enochs (1990) have modified Gibson and Dembo's instrument to measure teacher efficacy in a specific content domain or have added items to the instrument for specified reasons.

For example, Riggs and Enochs (1990) developed a measure of teacher efficacy and outcome expectations specifically for elementary science teaching. They modeled their items after Gibson and Dembo (1984). These modifications attempt to make the assessment of efficacy more specific but do not address the other problems of congruence with Bandura's theory discussed above. Thus, the instruments are subject to the same questions of validity addressed in relation to Gibson and Dembo's instrument.

Guskey's Responsibility for Student Achievement Questionnaire

Guskey and Passaro (1994) examined the construct of teacher efficacy adapting the research of Gibson and Dembo (1984) and Woolfolk and Hoy (1990). They believed the two factors of teaching efficacy and personal teaching efficacy were confounded with referent and locus of control. All personal teaching efficacy items were positively worded using "I" as the referent and an internal locus. Teaching efficacy items were negatively worded using "teachers" as the referent and an external locus.

To investigate this possible confounding, Guskey and Passaro developed a new instrument by randomly selecting 7 of the 12 personal teaching efficacy items from the Woolfolk and Hoy (1990) study, all of which reflected a personal-internal orientation, and rewording them to reflect a teaching-internal or a personal-external orientation. For example, the question, "When a student does better than usual, many times it is because I exert a little extra effort," was reworded to say, "When a student does better than usual, many times it is because the teacher exerts a little extra effort." In this case, the item was reworded to reflect a teaching-internal orientation. Four of the nine teaching efficacy items from the same study were also randomly selected and reworded to reflect a personal-external or a teaching-internal orientation. All items were then reassembled to construct the instrument.

Guskey and Passaro administered the instrument to 342 participants and performed factor analytic procedures and found a two-factor solution similar to the previous findings of Gibson and Dembo (1984) and Woolfolk and Hoy (1990). However, contrary to those findings, these factors corresponded not to a personal teaching efficacy versus teaching efficacy distinction, but instead to an internal versus external distinction similar to the locus-of-control construct identified in causal attribution theory. The internal factor appeared to represent teachers' perceptions of their and other teachers' ability to have impact on or influence in certain classroom situations. The external factor was related to influences on student behaviors that were outside the classroom. Teachers with high efficacy believed that even when faced with strong negative external factors (e. g., poor social, demographic and economic conditions), they could still have a powerful influence on their students. This study raises the question of what the Gibson and Dembo (1984) Teaching Efficacy Scale is assessing, but still does not address the incongruence with Bandura's expectancy construct.

Guskey (1981; 1987) used the construct of teacher efficacy to mean teachers' personal sense of responsibility for student success and failure. He developed and validated the Responsibility for Student Achievement Questionnaire (RSAQ). This questionnaire presents teachers with familiar negative and positive student achievement experiences and asks them to estimate the percentage to which the event was caused by an internal factor related to the teacher and the percentage to which the event was caused by external factors beyond teachers' immediate control.

The total percentage of the two factors should equal 100%. Smylie (1990) criticized Guskey's use of this definition and measure of teacher efficacy because he frames the sense of efficacy in terms of teachers' roles in past accomplishments. Smylie appropriately contended that Bandura's efficacy expectation construct should be framed in terms of teachers' perceived capacities for present or future behaviors or actions.

Denham and Michael (1981) suggested that teachers' causal attributions, as measured by Guskey above, are antecedent to teachers' efficacy and outcome expectancy judgments. For example, if a teacher attributes her failure to lack of ability, efficacy expectations may decrease. But, if she attributes her failure to students' background, then her efficacy expectations may be preserved. This process may also work in reverse with teachers' efficacy and outcome expectations influencing the attributions they make for successes and failures. If teachers have low outcome expectations, then they may attribute failure to external and/or unstable factors beyond their control in order to preserve efficacy expectations and self-esteem. If they have high outcome expectations they may blame themselves and their inabilities which would only serve to reduce efficacy expectations further. Teachers with high efficacy expectations might attribute success to internal, stable, and controllable factors such as their ability as teachers to employ techniques to overcome external factors that may hinder effectiveness. Teachers with high efficacy and/or outcome expectations might attribute their failures to internal, unstable factors that they can control the next time.

Global and Content Specific Teacher Efficacy

Curda, Curda, and Kleine (1996) developed an instrument to assess the generality and strength dimensions of teacher efficacy expectations through measurement of content specific teacher efficacy and global teacher efficacy. From Bandura's (1986) theory, they defined content specific efficacy as a teacher's judgment of his or her capability to successfully perform behaviors related to teaching a specific subject. More specifically, to what extent do teachers believe they possess skills to prepare materials, plan lessons, and motivate students in math, reading, writing, social studies, and science. Examples of their content specific efficacy items are, "I am confident I have the ability to apply cooperative group strategies when teaching Math" and "I am confident I have the ability to prepare engaging materials to teach Reading." Both the definition and statements reflect better alignment with Bandura's efficacy expectation construct than the ones discussed previously. First, the definition and statements are assessing the teacher's confidence about the link between the person (the teacher himself) and his ability to perform a specific behavior (apply cooperative group strategies or prepare engaging materials). Second, the statements are assessing generative capabilities rather than component acts. Both "applying cooperative group strategies" and "preparing engaging materials" require teachers to combine subskills into integrated courses of action to serve a variety of outcomes. Third, their assessment of teachers' efficacy across a variety of content domains allows determination of the generality of elementary teachers' efficacy expectations.

Global efficacy was defined as a teacher's judgment of his or her capability to carry out behaviors needed to insure student learning. Global efficacy is an assessment of teachers' confidence about whether or not they possess the skills and abilities to successfully perform behaviors related to managing a classroom, disciplining students, and meeting the learning needs of individual students. Examples of their global efficacy items are, "I am confident I have the skills to reinforce students for good behavior," and "I am confident I have the ability to maintain smoothness and momentum in my lessons." Once again, the definition and its corresponding statements reflect better alignment with Bandura's efficacy expectation construct. The statements are linking the person with a specific behavior, and the behaviors are generative capabilities not component acts. None of the statements were placed in the context of teaching a specific content area. This reflects the researchers' intentions to assess teachers' efficacy for performing behaviors not specifically connected to a content area but associated with activities in which teachers are likely to engage regardless of the content of the lesson. The most pronounced weakness of their study is that their instrument did not include assessment of Bandura's outcome expectancy construct in relation to teaching.

Results from a factor analysis suggested that teachers distinguish between math, science, reading-writing (collapsed to form a single subscale), social studies, and global efficacies. This measure attempted to investigate the generality of efficacy in elementary preservice teachers and found that teachers' efficacy expectations do vary across different content domains. The content specific nature of teacher efficacy as evidenced by these results also suggests that the measures of Gibson and Dembo and others reviewed above are inadequate due to the global nature of these items. Curda, Curda, and Kleine's (1996) definitions and operationalization of their teacher efficacy dimensions seem to be more clearly in alignment with

Bandura's concept of efficacy expectations and are likely to be more useful in identifying specific sources of teacher inefficacy for the purposes of development of interventions as Ashton (1985) suggested. However, this study did not develop or investigate measures related to Bandura's concept of outcome expectations.

From the above variety of definitions of teacher efficacy expectations and teacher outcome expectations, one can see that each researcher uses slightly different terms and words to define each construct which leads to clear distinctions in the operationalization of these constructs. This results in a variety of instruments proposed to measure one or more of the constructs and leads to some confusion. Each researcher cites Bandura's work and contends to be measuring teacher efficacy as he has proposed. The above critiques of these definitions and measures of teacher efficacy and outcome expectations calls into question the validity of these assertions.

The Current Project

In summary, a significant amount of research related to the definition and measurement of the constructs of teacher efficacy expectations and teacher outcome expectations exists. Unfortunately, this has not led to clear definitions of the constructs or development of a measure from which valid and reliable inferences can be made. Individual researchers continue to define teacher efficacy expectations and teacher outcome expectations in unique ways and to develop their own, borrow, or combine several measures in order to investigate factors influencing teacher efficacy expectations and teacher outcome expectations, and what influences they have on a variety of teacher and student motivations, teacher and student behaviors, and student outcomes.

A great deal of theory and model development has been generated from thinking about how teacher efficacy relates to teacher behavior and student motivation and achievement. We think we know the possible influences teacher efficacy expectations and teacher outcome expectations have on teacher behaviors and student motivation and achievement. However, few research studies have actually been conducted to test various models of the relationships that exist among these variables. The studies reviewed above and their significant findings lend support for the need for further investigation of this construct, how it can be measured, and its influences on variables such as teacher motivation and behaviors related to effort, persistence, and choice of instructional activities, and student motivation, cognitive engagement, and achievement. In addition, these findings have contributed to theoretical and empirical research attempting to identify key influences on teacher efficacy expectations and teacher outcome expectations in an effort to maximize those variables found to positively influence teacher efficacy expectations and teacher outcome expectations and minimize those variables that hinder them.

To further investigate various aspects of the constructs of teacher efficacy and outcome expectations, development of a measure of teacher efficacy and outcome expectations from which valid and reliable inferences can be made is crucial. The variety of more or less appropriate definitions and measures of teacher efficacy and outcome expectations reviewed previously in this paper suggests that the development of a measure in line with theory and the purposes of research is needed. Several researchers discussed this need and provided suggestions for development of efficacy measures. The following project used much of this advice to construct and conduct validation studies on a measure of teacher efficacy expectations and teacher outcome expectations in the content domains of reading and math.

CHAPTER III

Instrument Development Methodology

Gable and Wolf (1993) provide detailed suggestions for the development of instruments in the affective domain. The goal of their process of instrument development is to ensure that valid and reliable inferences can be made using the instrument. The steps included in this process are: (a) review of related literature, (b) development of conceptual and operational definitions of the construct of interest, (c) development of instrument specifications, (d) conducting content validity studies, (e) conducting a pilot of the instrument, (f) conducting construct and criterion-related validity studies, and (g) conducting reliability analyses. There are a variety of methods that can be used to complete each of these steps. Explanation of the details of the procedures I employed in this project are the focus of this section.

Instrument Specifications

The first step in the development of any instrument is a comprehensive review of related literature for the purpose of the development of conceptual definitions with a theoretical base. Evidence of the review I conducted is in the previous chapters of this dissertation. Next, one should develop operational definitions. These are the belief statements that can be judged to be either positive or negative in direction. The instrument developer then selects a scaling technique and response format for the belief statements. These procedures resulted in my development of the instrument specifications in Appendix A. These specifications included the conceptual definitions, operational definitions (represented by a sample belief statement for each conceptual category), and a description of the intended response format.

The instrument specifications indicate that separate groups of items are developed for judgments of efficacy expectations and outcome expectations in two content domains - reading and mathematics. Items were developed for these domains to investigate the generality or content-specific nature of teacher efficacy expectations and teacher outcome expectations. Outcome expectancy items were developed that correspond to the efficacy items in each domain. Each efficacy expectation and outcome expectation item references a particular instructional task or strategy in which a teacher may engage within the domains of reading or mathematics. Also, within each domain, the items developed intended to sample from the range of perceived difficulty/challenge of performing the instructional tasks and strategies.

Bandura suggested that efficacy expectations should be measured along a dimension of magnitude to investigate whether such expectations are limited only to simple tasks within the domain or whether they are inclusive of moderately challenging and difficult tasks. Items sampled from the range of difficulty aided in measuring the magnitude dimension of efficacy. The dimension of strength was measured through the response format of the percent certainty continuum.

The instrument specifications indicate the desirable final pool of items for each domain of efficacy is lower than the number of items presently used in the instrument. Instrument development procedures suggest to begin with more than the number of items you wish to have in your final pool. Most likely, some items will be dropped at each stage of the instrument development procedures. These instrument specifications were used to guide scale development. The initial instrument is in Appendix B.

Content Validity

After developing the conceptual and operational definitions one should conduct a judgmental review of the instrument items to address the issue of content validity. Content validation is one of the first steps in any instrument development. Content validation procedures assess the extent to which items on the instrument adequately sample from the intended universe of content (Gable & Wolf, 1993).

Content experts should be provided with a bibliography and summary of the literature used. They review the theoretical rationale and conceptual definitions that describe the universe of possible items for the instrument and judge the adequacy of each conceptual definition as it relates to the proposed use of the instrument using rating sheets. The same content experts also review the items and judge the correspondence between the conceptual definition and the items, usually through rating forms.

After any revisions are made from the feedback of the content experts, a second judgmental rating exercise can be conducted with a larger group of experts who are knowledgeable in the content of the instrument. These participants should be asked to read each item, assign it to the category it best fits, and indicate how comfortable they feel about that assignment (Gable & Wolf, 1993).

Preliminary Pilot of the Instrument

The next step after completing any revisions from results of content validity procedures is to prepare a draft of the instrument and gather preliminary pilot data. The purpose of this step in the validation process is to administer the instrument to a small representative sample and collect data regarding the clarity of the directions, readability, and the ease of responding (Gable & Wolf, 1993). Comments and reactions from this small sample can provide information that may be important for the success of your instrument. Initial item analysis can also be conducted during this step. Corrected item-scale correlations and Cronbach alpha internal-consistency reliability estimates can be computed for each of the hypothesized subscales. Investigation of correlations between items and their subscales and the estimated change in coefficient alpha reliability if items were dropped are used to make decisions about item inclusion and exclusion on the instrument for purposes of shortening the instrument for final administration or eliminating obviously poor items or those with ceiling or floor effects.

Construct Validity

In classical test theory, construct validation is the never-ending process of assessing the extent to which constructs explain covariation in the responses to the items on the instrument (Gable & Wolf, 1993). One can argue that an instrument actually measures a construct or constructs when relationships among the items proposed to measure that construct are judged to be consistent with the conceptual definitions. Empirically, the construct validity of an instrument is assessed through statistical procedures that analyze the extent to which items proposed to measure a construct share common variance. Exploratory and confirmatory factor analyses are often employed to provide empirical evidence of construct validity given this definition.

In contrast to classical test theory and in the context of latent trait theory, Wright and Stone (1988b) define construct validity as the relation between the difficulty order of the items produced by the way persons respond to these items and the item content. This relationship verifies or contradicts the intended definition and the meaningfulness of the variable. It assesses the extent to which items define the underlying continuum of interest (i. e., easy to difficult to endorse items for efficacy or outcome expectations) and if people are sufficiently dispersed along the same continuum. If these things are present, then it is possible to distinguish and describe high-and low-scoring individuals on the construct of interest. Construct validity can be defined and assessed by analysis of the expected relationship between the observed and expected responses. If the response patterns of items and persons fit the requirements of the measurement model then a construct has been well-defined and measured (Wright & Stone, 1988b). Wright and Masters (1982) stated that when an explicit measurement model is used, the internal validity of a scale can be analyzed in terms of the statistical fit of each item to the model that is independent of the sample distribution. If the fit statistics are acceptable (i. e., near their expected value) then you can say that the item calibration is valid. The internal consistency of each person's pattern of responses is analyzed in the same way, and if the fit statistics for persons' performance are acceptable then you can say that their measures are also valid. A variety of item response methods can be employed to provide empirical evidence of construct validity given this definition.

The methods of analyzing construct validity that I employed in this study are described in the following sections. First, I include a brief discussion of classical test theory and item response theory to provide the contexts from which I assessed construct validity. Second, I describe Rasch analysis procedures to illustrate how such procedures provide evidence for construct validity and reliability. Finally, I explain the utility of confirmatory factor analysis procedures to provide evidence for construct validity.

Classical Test Theory and Methods

Classical test theory is based on the conception of the observed test score as a composite of true and error components. The classical model is a model for the propensity distribution of an observed test score, meaning it considers the observed score for a fixed person and test as a random variable across replications (Hambleton & van der Linden, 1982). The model defines true score as a person's expected score over repeated administrations and error score as the difference between the true score and observed score. It has the primary assumption that the correlations between the errors on distinct measures are zero. The classical model is based on weak assumptions that can be met by most data sets so it has been applied to a wide variety of test development and test score analysis procedures. However, the classical model has practical difficulties and interpretation problems when it is applied to behavioral measurements (Hambleton & Swaminathan, 1985). Some of these problems are discussed below.

The classical model provides a test-dependent score meaning that every test entails its own true score even if a group of tests are known to measure the same ability or achievement variable. Therefore, in order to compare persons it is necessary for each of them to be given the exact same set of items. This is undesirable because what we would rather be able to do is have test-free person scores that do not depend on the particular items chosen for the test or the difficulty of those chosen items, but only give the positions of the persons on the variable of interest (Hambleton & van der Linden, 1982). This frees ability estimates from differences in test difficulty and allows (a) the possibility of persons taking different tests (e. g., parallel forms or a short and long form) measuring the same construct to receive measures of ability that are comparable, and (b) the ability to interpret this measure beyond the confines of the particular set of items on which the measure was based (Wright, 1996).

Another problem is that other item and test parameters, such as item difficulty, item discrimination, and reliability coefficients, are sample dependent. For example, test reliability is directly related to test score variability (Hambleton & Swaminathan, 1985). Thus, if the true score variance in a selected sample changes from a previous sample, and it most certainly will, then the reliability coefficient will also change. Therefore item and test statistics are often only useful in item selection when constructing tests for populations very similar to the sample on which the statistics were computed (Hambleton & van der Linden, 1982).

A third problem is the need for parallel measurements, which are difficult to obtain in practice, when using the classical model for obtaining reliability estimates and other parameters. People are never exactly the same during a second administration of a test yet classical test theory relies heavily on this methodology for its most important test and item parameters (Hambleton & van der Linden, 1982). Also, the classical model does not allow determination of the probability of success on a specific item given a specific person and his or her ability, and it assumes that the variance of errors is the same for all persons even though it is not uncommon to have people who perform more or less consistent than others (Hambleton & Swaminathan, 1985). Other limitations of classical test theory include the inability of its procedures to provide adequate solutions to the identification of biased items and the equating of test scores (Hambleton & Swaminathan, 1985).

Item Response / Latent Trait Theory

Item response theory (IRT) or latent trait theory models assume a different relationship of the test score to the variable measured by the test. An IRT model specifies a relationship between the observed test score of a person and the psychological construct (e. g., self-efficacy) assumed to underlie performance on the test. Thus, IRT assumes that a person's performance on a test can be explained or predicted by defining the characteristics associated with the psychological construct and estimating the person's score on these characteristics. Within this theory a variety of models can be defined and the appropriateness of any of the models can be investigated by conducting an appropriate goodness of fit test (Hambleton & Swaminathan, 1985).

The IRT model that seems most appropriate is defined prior to the scoring of the test. The different IRT models are differentiated by the number of parameters estimated for the items and the nature of the item characteristic curves (ICC). An ICC is a plot of the level of performance on some task against the ability measured by the set of tasks that contains it. Simply, an ICC is the nonlinear regression function of item score on the psychological construct measured by the test. If one assumes that items contained in a test only vary in their difficulty then the one-parameter model, the Rasch model, is the model of choice. The one-parameter model describes test items in terms of their difficulties, and the ICCs are nonintersecting curves that differ only by a translation along the ability scale. If one assumes that the items on a test are likely to vary in their difficulty and will also vary in their ability to discriminate between persons of high and low ability then the two-parameter model is used. The two-parameter model describes items both by their difficulties and discriminations and has ICCs that vary in slope and translations along the ability scale. The three-parameter model is chosen if one expects that the test items will vary in difficulty, vary in their ability to discriminate, and contain an element of guessing. This model describes item difficulty and discrimination and adds a chance level or pseudo-guessing parameter which creates ICCs that vary in slope, translation, and lower asymptote (Hambleton & Swaminathan, 1985; Weiss & Davison, 1981).

Once the model is chosen, you statistically estimate the parameters of the model using the data gathered and assess the fit to the item response model. The modeling is aimed at the item level and recognizes the fact that the data gathered in educational or psychological assessments are qualitative responses to test items (e. g., right/wrong, yes/no, strongly agree/agree/disagree/strongly disagree). Using specified measurement models, quantitative information is obtained from the qualitative data. Ability parameters are estimated using the person response patterns and represent the positions of persons on the construct measured. Item parameters are estimated using the item response patterns and represent the properties of the item affecting the respondent responses (e. g., the difficulty of the item). Thus, IRT models give quantitative item and ability parameters to explain qualitative item responses. These models do not explain the actual item responses but the probabilities of the responses and provide a distributional form for them. IRT models are mathematical models and their focus on statistical estimation rather than measurement plays a central role in addressing many of the limitations of classical test theory (Hambleton & van der Linden, 1982).

IRT models replace the true score of classical test theory with the latent parameter which is not indexed to the test. These test-free scores make it possible to estimate a person's ability on the same ability scale from any subset of items in the domain of items that have been fit to the model. This estimate will be an unbiased estimate of true ability and any variation in ability estimates obtained from different sets of items is due to measurement error only.

This estimation of person parameters also allows for sample independent item parameters to be estimated. The person parameter is used to remove the effect of person ability on the probability of success on an item (Hambleton & van der Linden, 1982). These ability estimates also have well-defined standard errors of estimate and are expressed as functions of ability so that the standard error is reported for each level of ability rather than for the test as a whole (McKinley, 1989). In addition, through statistical analyses of these model parameter estimates, the precision (reliability) of the test can be examined without replications (Hambleton & van der Linden, 1982).

The ability of IRT models to overcome the many limitations of the classical test theory model, however, is only evident when the model chosen "fits" the test data well. When there is reasonable fit, IRT procedures provide both invariant item statistics and ability estimates. Items and persons are placed on the ability scale in such a way that there is as close a relationship as possible between the expected person probability parameters and the actual probabilities of performance for persons at each ability level (Hambleton & Swaminathan, 1985).

The Rasch Family of Models

The Rasch family of models are used to construct variables and make measures from data. Rasch models are mathematical models that specify additivity and unidimensionality. Thus, all items measure a single construct and the measurement units are the same size (i. e., interval) across the continuum. The family of models was developed independently of other IRT models but can be viewed as an IRT model in which the item characteristic curve is a one-parameter logistic function. The first five Rasch models developed were Rasch's Dichotomous model, the Poisson Counts model, the Binomial Trials model, the Rating Scale model, and the Partial Credit model. The models share a common algebraic form, and the last four listed above can be thought of as simple extensions of the dichotomous model to other response formats (Wright & Masters, 1982). All five models share the possibility of sample-free item calibration and test-free person measurement. They predict the probable response of a person to an item, given the person's location on the underlying trait and the item's difficulty in relation to the underlying dimension being measured (O'Brien, 1992).

The probability of an individual giving a particular response to a question is estimated using two terms. The person parameter estimates the location of the person on the underlying construct which the questions were designed to assess. The rating scale step measures and individual item difficulty estimates are used to represent the transition (or step) within that question between one response and the next (e. g., the step from right to wrong, from strongly disagree to disagree, or from disagree to agree). Rasch analyses allow the computation of the probability of a response as a function of the person's estimated ability and the estimated difficulty associated with the step(s) between adjacent response alternatives associated with a question (McArthur, Cohen, & Schandler, 1991). Response category probability curves illustrate the points along the ability scale (indicated by the intersections of the curves) where the likelihood of responding using a particular response changes to the next step (i. e., the next adjacent response category). For example, in the context of the dichotomous model, when a person has more of the latent ability than an item requires, then the person's ability exceeds the item difficulty and the person's probability of success is greater than .5. The more the person's ability exceeds the item's difficulty, the greater the difference and the higher is the person's probability of success. However, when the item difficulty exceeds the person's latent ability and their difference is negative, then the person's probability of success is less than .5 (Wright, 1977).

The Rasch family of models is a special case of the three-parameter logistic model in which all items have equal discriminating power and guessing is assumed to be minimal. The model assumes that all items are equally effective in discriminating among respondents. The item and person parameters are estimated on the basis of statistics that make use of all relevant data available. A respondent's score contains all the information needed for estimating his or her ability, and the item difficulties may be estimated from a count of the number of respondents completing each step of an item with polytomous categories (De Ayala, 1993). For example, a response of "agree" using a four point Likert-type scale would indicate a respondent has completed two steps, one step by deciding to respond with disagree rather than strongly disagree and another step by deciding to respond with agree rather than disagree. The steps made when responding to an item may not be equally difficult nor be ordered in difficulty. When the steps are not ordered in terms of difficulty a reversal is said to exist indicating that a step higher on the response scale continuum was more probable than a step lower on the continuum. In such cases, a response scale category will never be the most probable response. When the model is applied to a set of Likert items that share a fixed set of rating points it is expected that the relative difficulties of the categories for each item should be relatively similar across items (De Ayala, 1993).

To achieve the separation of person and item parameters in order to make objective comparisons, Rasch models do not use a slope parameter. The models are based on logistic item operating curves with the same slope. By modeling operating curves to have the same slope all person parameters and item parameters are point locations on a single latent variable so they can be expressed in the same scale units (Wright & Masters, 1982). The unit is called a "logit". Logits typically range from -4 to +4 with increasing person estimates indicating higher ability and increasing item estimates indicating greater difficulty (Ludlow, Haley, & Gans, 1992). For the dichotomous model, a person's ability in logits is their natural log odds for success on items defining the scale origin. An item's difficulty in logits is the natural log odds for failure on that item by person's with abilities at the scale origin (Wright, 1977). The item characteristic curve illustrates the way the probability of success on that item changes as a function of person ability when item difficulty is held constant and person ability is varied. The person characteristic curve illustrates the way a particular person is expected to perform on items of various difficulties when person ability is held constant and item difficulty is varied (Wright, 1977). Since the models include fewer parameters it is easier to work with and interpret. Problems with parameter estimation are considerably fewer in number than for the more general IRT models. For example, when an item is twice as easy as another then a person's likelihood of success on that item is twice that of the harder item. Also, if a person's ability is twice that of another then the person's likelihood of success is twice that of the second person. Other item response models do not permit this particular kind of interpretation of item and ability parameters (Hambleton & Swaminathan, 1985).

Often mentioned weaknesses of Rasch models are that the assumptions that all item discriminations are equal and guessing is minimal are restrictive and evidence is available to suggest that unless test items are specifically chosen to have these characteristics then the assumptions will be violated. Hambleton and Swaminathan (1985) suggested that both assumptions may be very doubtful for achievement test data. However, Forsyth, Saisangjan, and Gilmer (1981) investigated these invariance properties using achievement tests and found that Rasch models are robust even when such assumptions are not met. Wright (1996) also points out that one can construct a test in which guessing plays a big part or in which items vary widely in their discrimination, but he does not understand why anyone would want to do this if they aspire to objective mental measurements. When item discrimination is allowed to be an active parameter in the measurement model then person-free test calibration is unattainable (Wright, 1996).

The Rating Scale Model

The Rasch Rating Scale Model (RSM) was used in this project to provide evidence for validity, as well as for reliability. The RSM estimates person abilities and item difficulties for responses scored using at least two or more ordered categories (Wright & Masters, 1982). When this model is applied to the analysis of a rating scale, a position on the variable is estimated for each person (person ability), a scale value is estimated for each item (item difficulty), and step measures (response thresholds) are estimated for the rating categories (Wright & Masters, 1982). For an item with a response scale from one to four (strongly disagree, disagree, agree, strongly agree), a person's "step" can be thought of as choosing the *k* alternative over the *k-1* alternative. So, a person who chooses "agree" given these four categories has taken two steps, one when choosing "disagree" over "strongly disagree" and the next one when choosing "agree" over "disagree". This person failed to take the final step, choosing "strongly agree" over "agree" (Wright & Masters, 1982).

The relative difficulty of making these "steps" in a rating scale item is governed by a fixed set of rating points for all items. Because the same rating scale points are used for each item, the model assumes the relative difficulties of the steps in each item should not vary from item to item (Wright & Masters, 1982). Thus, the only difference remaining between items is their location on the latent variable (i. e., the difference in the overall difficulty of the items). The pattern of the *k* steps in each item relative to that item's difficulty is described by the step measures or threshold parameters and is estimated once for the entire set of items. A rating scale that operates in the manner intended will have ordered steps. Unordered steps would indicate that it was easier for respondents to make a higher step decision than a lower step decision. For example, unordered steps with the four category rating scale above might indicate that it was easier for respondents to go from agree to strongly agree than to go from disagree to agree once they passed the disagree category (Zhu, Updyke, & Lewandowski, 1997). Unordered categories may indicate poor category wording or the need to combine or omit categories.

One part of rating scale development is determining how many Likert categories should be used. RSM procedures can also be used to determine optimal categorization of Likert rating scales. Once optimal categorization is determined then additional characteristics of the optimal rating scale can be investigated, such as the rating scale's adequacy at discriminating different ability levels of persons or different difficulty levels of items. To begin analyzing the data all possible category collapsings are constructed through the recoding of the original category response data. Then each of the rating scale data sets are analyzed. Item and person Infit and Outfit statistics of all possible rating scales are first examined. Second, the average measures and step calibration measures are examined for each rating scale to determine the categorization order. The rating scale must have ascending order - a basic property of the categorization in a rating scale- on both of these sets of category statistics to be considered satisfactory. Third, the remaining rating scales are examined using the person and item separation and reliability statistics to determine the rating scale with the optimal discrimination of people and items (Zhu, Updyke, & Lewandowski, 1997).

Concepts of Reliability and Validity and Use of Person and Item Fit Statistics in Rasch Modeling

One purpose served by Rasch analysis is to estimate item difficulty and person ability. It assesses the extent to which items define the underlying continuum of interest (i. e., easy to difficult to endorse items for efficacy or outcome expectations) and if people are sufficiently dispersed along the same continuum.

Analysis of participant responses to the instrument items identifies if responses are distributed across the Likert or other response continuum. If these things are present, then it is possible to distinguish and describe high-and low-scoring individuals on the construct of interest. For example, one parameter in Rasch analysis will estimate a scale value (item difficulty) for each item. These measures indicate how easy it is for a person to agree to an item. An adequate sampling of items from the domain should result in items being distributed across the continuum (i. e., those easy to agree with and those difficult to agree with). In this case, Rasch provides another confirmation of content and construct validity because the ordering should correspond to theory or logic (Wright & Stone, 1988b).

Rasch analysis procedures also provide estimates of person parameters by estimating a position for each person on the continuum (i. e., efficacy and outcome expectations in each domain). Once again, if we have appropriately sampled from the population then we would expect different people to fall in positions along the continuum ranging from low to high efficacy or outcome expectations. If this does not happen and people all fall in the same relative location, then these items are not contributing to our ability to distinguish between individuals with high and low efficacy or outcome expectations or there was poor sampling of persons. Person statistics also have well-defined standard errors of estimate expressed as functions of ability. So, a standard error of estimate is reported for each level of ability, rather than for a test as a whole (McKinley, 1989).

The standard errors associated with each item difficulty and person ability estimate provide evidence for reliability (referred to as precision in IRT), defined as the degree to which test scores are free from measurement errors. These errors are used to describe the range within which each item's "true" difficulty or each person's "true" ability is located. These individual person ability errors can be squared and summed to produce a correct average error variance for the sample that is used to estimate reliability that will have an equivalent interpretation as the traditional coefficient alpha often used as an estimate of reliability (Wright & Stone, 1988a). The item and person separation indices provided by the Rasch analysis give measures of the extent of dispersion along the underlying continuum. The separation index is the ratio of the unbiased estimate of the sample standard deviation to the root mean square measurement error of the sample (Wright & Stone, 1988a). Reliability coefficients can be estimated using $G^2 / (1 + G^2)$ where G represents the separation index. Wright (1994) suggested that values greater than .90 indicate a clearly unidimensional variable measured by internally consistent items. Item and person strata, calculated as four times the separation index plus one divided by three [(4G+1)/3], indicate regions of the scale whose centers are separated by logit distances greater than can be explained by measurement error (i. e., beyond three standard errors). Kilgore, Fisher, Silverstein, Harley, and Harvey (1993) suggest that a scale must reach out to at least two item difficulty strata to be useful for scale definition.

The location of each item and person on a line representing the construct along with their standard errors shows the definition and utility of the construct. Evidence for the definition of the construct is provided by the item locations. The items must be well dispersed in difficulty to identify the direction and meaning of the construct. Evidence for the utility of the items for measuring persons is provided by the standard error for each person measure. To be useful, a measure must be able to separate relevant persons by their performance or responses. These item and person separation indices in Rasch provide a way to evaluate the successful development of a construct and its utility (Stone & Wright, 1988).

Rasch latent trait analysis also assesses how well the items and people fit the model based on the observed response pattern in relation to the expected response pattern specified by the model. The conformity of any set of items and any sample of persons to the Rasch model, and even the conformity of any particular item or person can be evaluated by fitting the Rasch model to the data, calculating the residuals from the values expected in the model, and examining these residuals (Wright, 1977). Individual item and person fit indices are more acceptable than global fit statistics for items and persons when investigating fit. The fit of the data to the Rasch model is evaluated by calculating how much is "left over" after the data have been used to estimate item difficulties and person abilities. The standardized square of this residual can be squared and summed over persons or items to form approximate chi-square distributed variables for testing the fit of any particular item to any group of persons, or of any individual person to a set of items (Wright, 1977). Fit statistics are derived from a comparison of the expected patterns and the observed patterns of response. These fit statistics are used as an assessment of the validity of the model-data fit and as a diagnosis for idiosyncratic item or person performance (Lusardi & Smith, 1997). Since the focus of this project is instrument development, the primary focus will be given to item fit statistics rather than person fit statistics.

An item's parameter estimate is accompanied by an error term and a fit statistic that allow the researcher to determine how well or poorly the item matches the underlying continuum (McArthur, Cohen, & Schandler, 1991). Once the item parameter is estimated, item fit statistics are used to verify the internal consistency of the items in contributing to a unidimensional scale. The model specifies that an item should have a greater probability of receiving a higher rating for a person with higher ability than for persons of lower ability (Lusardi & Smith, 1997; Smith, 1986; Wright & Stone, 1988b). Rasch analysis also provides a parallel determination of how persons are performing and which persons are acting idiosyncratically by computing a person parameter estimate, error, and fit for each person (McArthur, Cohen, & Schandler, 1991). The model specifies that a person of a given ability should have a greater probability of providing a higher rating on easier items than on more difficult items (Lusardi & Smith, 1997; Smith, 1986; Wright & Stone, 1988b).

There are two types of chi-square fit statistics for items and persons, Infit and Outfit. Infit statistics are sensitive to unexpected responses to items near a person's ability level. Outfit statistics are sensitive to unexpected responses on items far from a person's ability level (Linacre & Wright, 1994). When reported as mean squares these statistics have an expected value of one and a range from 0 to infinity. When reported as standardized versions of these statistics, they have an expected value of zero and a standard deviation of one. In both cases, values greater than the expected values indicate excessive variability while values less than the expected values suggest a lack of variability (Linacre & Wright, 1994; Lusardi & Smith, 1997). Wright & Linacre (1994) suggest that an acceptable range for both types of fit statistics reported as mean squares is .6 to 1.4. A reasonable range for both types of statistics reported as standardized values is -2 to +2. Items that fall outside these ranges should be targeted for possible revision or elimination. A person outside these ranges may not be from the target sample or the instrument content may not be appropriate for him or her (Lusardi & Smith, 1997).

Item fit statistics are used in several ways. First, an individual item fit statistic indicates the extent to which the use of that specific item is consistent with the way people have responded to the other items (Wright & Stone, 1988b). The global item fit statistic is used in the same manner as an estimate of the overall internal consistency of the items. Second, item fit statistics are used to identify items for which the actual responses differ significantly from the responses expected under the model. These fit statistics identify items to which the responses indicate the existence of a construct other than that shared by remaining items or the existence of item flaws. These misfitting items provide the opportunity for item editing or for reconsideration of the initial test development specifications (O'Brien, 1992). Any item can also be analyzed for bias with respect to the sex or culture of persons by calculating a regression of its residuals on indicators of these background variables. An item is considered biased when its difficulty is different for one group than for another (Adams & Wright, 1994). Data from items that are found to be biased can be deleted from persons' responses without interfering with estimates of person ability. Thus, one can correct for item bias without losing the information available from the unbiased items (Wright, 1977). Investigation of these residuals makes the statistical detection of item bias possible and provides an objective quantitative basis for eliminating or correcting it (Wright, 1977).

Person fit statistics are used to identify individuals whose actual responses differ significantly from the responses expected under the model. Misfitting persons indicate they experience the variable differently than its usual manifestation (O'Brien, 1992). People whose responses are inconsistent with the item difficulty ordering shared by the majority of others can be identified for further investigation using the content of their response pattern, response frequencies or residual analysis. A more extensive analysis of the response pattern of each person involves evaluating the correlation of their residuals with item difficulty, position, and type. The standardized residuals are used in their unsquared form so that the distribution is centered at their expected mean and scaled by their expected standard deviation, and the distribution is approximately normal and their expected error variance is one (Wright, 1977). If, for example, guessing or lack of speed influences persons' responses, then regressing their response residuals on item difficulty will bring that out. This may help in identifying any unintentional item bias or ambiguity in the instrument for a particular subset of the sample (e. g., it may be that all teachers from a rural school setting have a different response pattern on specific efficacy expectation items than teachers in other settings). These statistics along with others provided through Rasch modeling analyze person characteristics and item structure that assist in providing finer score interpretations.

The use of fit statistics has also been proposed for evaluating the dimensionality of data. Smith and Miao (1994) found in a simulation study that the mean-square outfit statistic accurately identified multidimensionality. Smith (1996) and Smith and Miao (1994) also used principal components analysis to identify multidimensionality and found that it works well with rating scale data. Both concluded that researchers should use both methods in tandem. Smith (1996) and Smith and Miao (1994) suggested that where the intention is to create a unidimensional structure, one would expect few items to load on a second factor and the correlation between the factors to be high. In such situations, the item fit approach using Rasch detects dimensionality more accurately. Rasch analysis will be used as the primary analysis for validity and reliability purposes. The resulting item pool from the Rasch results will be submitted to a confirmatory factor analysis.

Confirmatory Factor Analysis

The purpose of factor analysis is to examine the interrelationships among instrument items and verify clusters of items that share sufficient variation to justify their existence as a construct being measured by the instrument. Confirmatory factor analysis (CFA) allows specification of a unique factorial solution, statistically evaluates the fit of the model to the data, and suggests possible modifications for model improvement (Mueller, 1996). In CFA, a model which the data are expected to fit is proposed that identifies the number of factors to be derived and which items are related to each factor. The results indicate how well the empirical data actually fit the proposed model.

Tabachnick & Fidell (1996) suggest multiple indicators should be used for examination of model-data fit. The most often used indicators include (a) the chi-square (χ^2) to degrees of freedom (df) ratio; (b) the comparative fit index (CFI, Bentler, 1990); (c) the goodness of fit index (GFI, Jöreskog & Sörbom, 1988); and (d) the adjusted goodness of fit index (AGFI, Jöreskog & Sörbom, 1988). With large sample sizes, the χ^2 statistic is sensitive to trivial differences between the sample and estimated population covariance matrices and should be interpreted carefully (Mueller, 1996). Using the χ^2 to df ratio, the rule of thumb for a good fitting model ranges from a liberal ratio of less than 5:1 (Gable & Wolf, 1993) to a conservative ratio of less than 2:1 (Tabachnick & Fidell, 1996). The CFI is a comparison of the specified model with a null model in which each manifest variable represents a factor. The GFI is comparable to $\underline{R}^2$ values in multiple regression and represents a measure of the relative amount of variance-covariance in the observed data that is accounted for by the specified model covariance matrix. The AGFI is an adjusted GFI that takes into account the number of parameters estimated in the model (Mueller, 1996). The CFI, GFI, and AGFI indices range from 0 to 1.00 with values greater than .90 indicating acceptable fit (Mueller, 1996).

If fit indices indicate good model-data fit, it is important to remember that this does not imply that the hypothesized model is the true underlying structure of the data. An analysis of the same data with a different hypothesized model could yield the same results (Mueller, 1996). If any of the fit indices indicate poor model-data fit, it is possible to locate internal specification errors using modification indices (MI) with associated expected parameter change statistics (EPC). For each parameter that is fixed in the specified model, the MI is an estimate of the decrease in the χ^2 badness of fit measure if the parameter were freed, and the EPC is the estimated value of that freed parameter (Mueller, 1996). Lagrange Multiplier (LM) statistics are modification indices useful for judging the worth of including more free parameters in the model to improve overall model-data fit (Mueller, 1996). LM statistics provide the researcher with suggestions for improving model-data fit. The researcher can choose to modify the structure as suggested and reanalyze and reevaluate the model-data fit. Such modifications should be theoretically justifiable and based on logic. In such cases, the fit results usually will improve, not necessarily due to a truly "better" model but simply because a model has been fitted to a particular sample data set (Mueller, 1996).

Criterion-related Validity

Criterion-related validity is concerned with the extent to which scores on an instrument are related to other external criteria (Gable & Wolf, 1993). Empirical evidence for criterion-related validity most often involves correlational analyses between scores on the instrument of interest and scores from other measures, such as observational reports or measures of behavior. Many feel that criterion-related validity also provides further support for construct validity.

Reliability

Reliability refers to the degree to which tests scores are free from measurement error and concerns both the internal consistency of the subscale scores and the stability of scores over time. As discussed previously in the Rasch analysis section, results from the Rasch Rating Scale analysis provides evidence for reliability. The internal consistency of the items is indicated by the individual item fit statistics. The standard errors associated with the person ability measures tells how precisely we are able to measure each person's ability, regardless of the sample drawn from the target population to which he or she belongs, when the items are internally consistent. The item and person reliability measures provided by Rasch results additionally indicate the extent to which items and persons are spread out along the continuum of interest for purposes of distinguishing easy and difficult items and persons with high and low efficacy or outcome expectations. A measure of stability does not seem to be warranted for an instrument such as the one developed for this project where individual teachers' efficacy and outcome expectations may differ according to the students in their present class. Also a repeated administration of the instrument is an impossibility due to the costs of conducting a second large mail survey.

CHAPTER IV

Validation Studies

This chapter focuses on the methods, results, and discussion of the individual validation studies that made up this project. The participants, instruments used, and results of each study, as well as the related discussion are reported successively because of the way in which each of the results inform the instrumentation and procedures used in the study to follow it. The project was broken down into eight studies for purposes of clarity in reporting. Four of the studies fall under the heading of content validation studies. The final four studies consist of reports on the final administration of the instruments for each content domain and fall under the heading of studies assessing construct validity, criterion-related validity and reliability.

Content Validation Studies

The following four studies were conducted after the initial instrument was constructed using the item specifications in Appendix A. The goal of the first study was for content experts to verify the theoretical rationale behind the conceptual definitions and the correspondence between the conceptual and operational definitions. The goal of the second study was for content experts to examine the extent to which the items truly reflect the categories and the domains of interest. The goal of the third study was to verify that the content of the items sampled across the continuum of easy to difficult instructional tasks. The goal of the fourth study was to conduct a pilot of the instrument to examine the clarity of the directions and the content of the items prior to conducting a final administration for purposes of performing construct and criterion-related validity and reliability procedures.

Study 1: Content Validity - Conceptual and Operational Definitions

Participants

Three professors and two graduate students, four male and one female, were asked and agreed to participate as content experts. One professor has published and is currently engaged in research related to teacher efficacy. The other two professors have current research interests in motivation. One professor's research is specifically concerned with the measurement of efficacy and the validation of such instruments; the other professor is currently engaged in research relating motivational variables, including efficacy, to cognitive engagement and achievement in the classroom. The two graduate students are presently engaged in their dissertation researches and have current research interests in motivational variables. Both have had advanced coursework in motivation research and measurement.

Instrument and Procedures

The five content experts were given a review of the theoretical rationale and conceptual definitions of teacher efficacy expectations and teacher outcome expectations in relation to the domains of interest (see Appendix C) and a copy of the present instrument (see Appendix B). They were also given a rating form for assessing the adequacy of the conceptual definitions as they relate to the proposed use of the instrument and assessing the correspondence between the conceptual definitions and the instrument items. The rating form can be viewed in Appendix C. Each participant received a copy of the rating form and related information. Each of them completed the rating form independently and within a time period of one week. Any necessary changes to the instrument were made prior to the next step of the content validation.

Results

The data from each form were compiled and can be viewed in Table 1. The table contains the individual ratings of each content expert. The ratings received for the items ranged from four to six on the six-point scale indicating all content experts believed the conceptual and operational definitions were above average in clarity and correspondence to theory. All feedback given in the form of comments was assessed and I respond to each one in the following discussion section.

Discussion

The results of this content expert analysis of the conceptual and operational definitions used for the development of the instrument were satisfactory. The high ratings on each of the items were encouraging. Several comments were given that made it necessary to consider making adjustments in the instrument prior to the next stages of content validation; other comments were simply acknowledged.

One comment made in relation to the conceptual definitions was "I am concerned about the recommended task item descriptions. I am not sure if easy to difficult should be the only focus. Are the tasks authentic reading instruction tasks? Have the teachers been instructed in how to use the various tasks? Who determined the difficulty of the tasks?" I acknowledged this first comment about the concern of the task item descriptions, and responded with the following explanation. Easy to difficult is one of the foci, but not the only one. The reason for this focus in particular is because Bandura's theory posits several dimensions of efficacy expectations. These are generality, magnitude, and strength. This instrument

Table 1

Content Expert Results

Item	Individual Ratings				
1. Conceptual definition of reading efficacy expectations	4	5	5	6	6
2. Conceptual definition of math efficacy expectations	4	5	5	6	6
3. Correspondence between reading efficacy definition and items	5	5	6	6	5
4. Correspondence between math efficacy definition and items	5	5	6	6	5
5. Conceptual definition of reading outcome expectations	5	5	5	6	6
6. Conceptual definition of math outcome expectations	5	5	5	6	6
7. Correspondence between reading outcome definition and items	5	5	6	6	6
8. Correspondence between math outcome definitions and items	5	5	6	6	6
9. Adequacy of definitions as they relate to the proposed use	5	5	6	6	6
10. Adequacy of the items as they relate to proposed use	5	5	6	6	6

measures generality across the domains of reading and math, and proposes to capture magnitude through the inclusion of easy, moderate, and difficult items. The difficulty of these tasks will be determined by a sample of elementary school teachers in a later content validity study. While I expect that teachers will have had instruction in how to implement many of the instructional activities, I am not assessing their preparation for these tasks. The assessment of efficacy is concerned with teachers' confidence that they can successfully perform these tasks. If teachers have received no instruction on how to implement the various strategies, I would expect their efficacy to be rather low. However, there may be other reasons for their low efficaciousness; and, in the present study, I am not concerned with the attributions for teachers' low or high efficacy.

There were two comments related to the correspondence between the efficacy definitions and the items. The first comment was, "The question about disabilities (ID, bilingual) assumes that you would use the same techniques with all of these students - that is probably not what a "good" teacher would do." My response was that I do not think this question implies that a teacher should or would use the same techniques with all of these students. I believe the use of the word "appropriate" before "learning experiences" should signal to teachers that a "good" teacher finds experiences that are appropriate for each learner. This would rarely entail using the same technique with all students.

The second comment was, "The questions about computers assumes that the school has the resources for technology. What if the school doesn't have the funding/equipment? I guess my concern is about teachers who feel highly efficacious but don't have the technology in their schools ---as you mentioned, you may need to look at these items separately." I agree that some teachers may not have the resources for technology in their school or classroom. However, my focus is not whether they have the technology, but, if they did have it, how efficacious they would be. I think I probably need to add a sentence or two to the directions related to the technology questions, something to the effect of, "While I realize that not all teachers have access to the following technologies, I would like you to report your confidence in using these technologies for the purposes described regardless of whether or not you actually have access to them in your school."

Additional comments were made about the correspondence between the outcome definitions and items. One comment was, "Your scale (using percentages) might be difficult for some teachers...can a teacher reliably distinguish between 39% and 40%?" This expert's concern about the response scale was also one of mine. The ranges may be too wide or may not be distinct enough. I decided to reserve further judgment of the adequacy of the response scale until after the pilot of the instrument. Other comments were, "Items 17 and 24 are not as specific as the other outcome items for reading," and "Items 16 (for reading) and 40 (for math) are more like efficacy items." I decided not to reword item seventeen or twenty-four until I piloted the instrument. I agree that item seventeen may not be as specific as some of the other items, but item twenty-four gives teachers a specific task they are to do using the World Wide Web. The concern over items sixteen and forty was warranted. The corresponding efficacy items did not translate well into outcome expectation items. I decided to reword these. The reworded item was the statement, "I am confident that an effective strategy for improving my students' standardized reading (or math) achievement scores is integrating well-designed computer software." Final comments included, "Instrument items fit the definition well," and "Looks good. I like it. It's about time we start analyzing specific tasks."

In summary, few adjustments were made to the instrument items prior to continuing with the content validity procedures and pilot of the instrument. At this point, the content experts agreed that the conceptual definition and operational definitions for teacher efficacy expectations and teacher outcome expectations in the content domains of reading and mathematics are theoretically sound and correspond well with one another.

Study 2: Content Validity - Item Categorization

Participants

Twelve graduate students enrolled in courses in a College of Education at a university in the midwest participated in this study. There were seven females and five males. Four of the twelve participants had previously or were presently teaching at the elementary school level. All of the participants had completed coursework in human learning and motivation and were knowledgeable of Bandura's theory of efficacy and outcome expectations.

Instrument and Procedures

The rating form (see Appendix D) asked participants to read each instrument item and indicate whether they believed it reflected a statement of a math or reading efficacy or outcome expectation and to report how strongly they felt that their decision was correct. They were asked to provide a rating from "1" (absolutely sure) to "4" (not very sure) indicating how strongly they felt about their placement of the item. This rating form allowed me to examine the extent to which the items truly reflect the categories and the domains of interest.

Each participant received the directions and rating form in Appendix D and completed the rating form independently within a time period of one week. For each item, the frequency and percentage of assignment to each category was calculated, and a criterion of 90% agreement was necessary for an item to remain in the category without revision. For those items that remained in each category a mean confidence rating was calculated using participants' ratings of how strongly they felt their decision was correct. This information was used as additional evidence when identifying items needing no revision, items needing revision, or items that may be dropped.

Results

The twelve participants completing the item categorization rating form had 100% agreement on the placement of items in each category and were in alignment with the original construction of the items. In addition, all participants indicated they were "very sure" or "absolutely sure" about their placement of each of the items. The mean confidence ratings for the items ranged from 3.4 to a perfect 4.0.

Discussion

The criterion of 90% agreement necessary for items to remain in the instrument without revision was met for all items. All participants reported high confidence in their categorization of the items. These findings provide further evidence for the content validity of the instrument. As a result of this analysis, no items were deleted or altered prior to the pilot administration of the instrument to a sample of practicing teachers.

Study 3: Content Validity - Item Difficulty Rating

Participants

Eleven elementary school teachers currently teaching in schools in the midwest participated in this content validity study. All participants were female. Three participants were teaching first grade, two were teaching second, one was teaching third, four were teaching fourth, and one was teaching fifth grade.

Instrument and Procedures

The rating form (see Appendix E) asked participants to rate the instructional activities for each content domain on the degree of difficulty the average teacher might have implementing it effectively. This rating form also asked participants to rate the adequacy of the sampling from all possible instructional activities and to identify any additional instructional activities they believed would improve student achievement that were not listed.

Each participant was asked to complete the item difficulty rating form in Appendix E. The participants completed the form independently within the time period of one week. The average difficulty rating of each set of instructional activities was examined to investigate whether there was a range of easy, moderate, and difficult tasks included in the instrument. The list of additional instructional activities was compiled with frequency counts when needed. If I needed to add items, instructional activities from this list were used if possible. Also, if there was a high frequency of a particular instructional activity it was considered for inclusion as an item on the instrument.

Results

The mean difficulty ratings for the instructional tasks for reading are provided in Table 2. The mean difficulty ratings for the instructional tasks for mathematics are presented in Table 3. Each set of tasks is arranged from easy to difficult.

For the reading tasks, the means ranged from 2.09 to 4.45. These participants reported that applying cooperative group strategies, integrating writing activities, and conducting whole group instruction are relatively easy instructional tasks to implement during reading. They reported that organizing activities around students' background, integrating the use of e-mail, and integrating multimedia projects were among the most difficult tasks to implement during reading instruction. The mean rating of the representativeness of the instructional tasks for reading was 5.2.

For the math tasks, the means ranged from 2.09 to 4.18. These participants reported that applying cooperative group strategies, providing worked examples, and conducting whole group instruction are relatively easy instructional tasks to implement during math. They reported that organizing activities around students' background, integrating the use of spreadsheets, and integrating multimedia projects were among the most difficult tasks to implement during math instruction. The mean rating of the representativeness of the instructional tasks for math was 5.5. None of the participants provided suggestions of additional instructional activities.

Discussion

As I expected the majority of the technology-related tasks fell in the upper end of the distribution indicating most teachers rated these tasks as more difficult to implement. This was the case in both sets of instructional activities. The easiest

Table 2

Means, standard deviations, and ranges of item difficulty ratings for instructional activities in reading

Instructional Activity	M	SD	Range
Applying cooperative group learning strategies in Reading	2.1	1.0	1 - 4
Integrating writing activities into Reading instruction	2.3	.8	1 - 4
Conducting whole group instruction in Reading	2.3	1.2	1 - 5
Integrating Reading activities into other curriculum areas	2.4	1.1	1 - 4
Providing instruction in phonics skills when teaching Reading	2.5	1.1	1 - 5
Planning effective lessons in Reading	2.6	.9	1 - 4
Using drill and practice software to reach instructional goals in Reading	2.8	1.0	2 - 5
Instruction in Reading based on students' interests	2.8	1.1	1 - 4
Conducting small group instruction in Reading	2.8	1.3	1 - 5
Providing students with the opportunity to choose their own literature books for Reading instruction	3.0	1.3	1 - 5
Preparing engaging materials in Reading	3.2	.6	2 - 4
Incorporating word processing activities into Reading instruction	3.2	1.0	2 - 4
Providing challenging seatwork in Reading	3.4	.8	2 - 4
Provide appropriate Reading learning activities for diverse learners	3.4	.9	1 - 4
Selecting appropriate computer software to integrate into Reading instruction	3.5	.8	2 - 5

(table continues)

Instructional Activity	M	SD	Range
Using computer tutorials to provide individualized instruction in Reading	3.5	1.0	2 - 5
Providing activities that allow students to use technology to attain an instructional goal in Reading	3.6	.8	2 - 5
Adjusting activities in Reading to account for differences in individual students' skills	3.6	.9	2 - 5
Adjusting activities in Reading to the learning needs of individual students	3.7	1.1	1 - 5
Using CD-ROMs that contain children's literature books that match the goals of the Reading curriculum	3.8	1.2	2 - 5
Integrating the use of e-mail into relevant Reading activities	3.9	1.3	1 - 5
Organizing activities in Reading according to students' background	4.2	1.0	2 - 5
Planning Reading lessons that allow students to use the World Wide Web to access relevant sources for curriculum related projects	4.3	1.3	1 - 5
Planning activities for students to author multimedia projects in Reading (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage)	4.5	.8	3 - 5

Table 3

Means, standard deviations, and ranges of item difficulty ratings for instructional activities in math

Instructional Activity	M	SD	Range
Providing worked examples during Math instruction	2.1	.8	1 - 4
Conducting whole group instruction in Math	2.1	1.1	1 - 5
Applying cooperative group learning strategies in Math	2.2	1.1	1 - 4
Providing Math instruction using a variety of hands-on manipulatives	2.2	1.3	1 - 4
Providing times practice tests for student during Math instruction	2.2	1.3	1 - 5
Using drill and practice software to reach instructional goals in Math	2.4	1.1	1 - 5
Planning effective lessons in Math	2.5	.9	1 - 4
Integrating Math activities into other curriculum areas	2.5	1.0	1 - 4
Providing challenging seatwork in Math	2.6	.9	1 - 4
Conducting small group instruction in Math	2.6	1.1	1 - 5
Preparing engaging materials in Math	2.8	1.0	1 - 4
Providing activities that allow students to use technology to attain an instructional goal in Math	2.9	1.3	2 - 5
Adjusting activities in Math to the learning needs of individual students	3.0	.9	1 - 4
Using computer games that match the goals of the Math curriculum	3.0	1.1	1 - 5
Adjusting activities in Math to account for differences in students' skills	3.1	.7	2 - 4

(table continues)

Instructional Activity	M	SD	Range
Using computer tutorials to provide individualized instruction in Math	3.1	1.2	2 - 5
Selecting appropriate computer software to integrate into Math instruction	3.1	1.4	2 - 5
Provide appropriate Math learning activities for diverse learners	3.2	.8	2 - 4
Instruction in Math based on students' interests	3.4	1.4	1 - 5
Organizing activities in Math according to students' background	3.6	.9	2 - 5
Incorporating spreadsheet activities into Math instruction	3.6	1.3	1 - 5
Planning Math lessons that allow students to use the World Wide Web to access relevant sources for curriculum related projects	3.9	1.4	1 - 5
Planning lessons that use videodiscs for instructional goals in Math	4.0	1.2	2 - 5
Planning activities for students to author multimedia projects in Math (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage)	4.2	1.0	2 - 5

and most difficult tasks were similar for both the reading and math domains. Overall, teachers reported that the instructional tasks listed were easier to implement in the context of teaching math than in the context of teaching reading. The ranges and standard deviations indicated that participants used the complete range of the response scale on many of the items and varied somewhat in their opinions of the ease or difficulty of the tasks. The participants also reported that the tasks included on each of the lists were a good to excellent representation of the range of teachers' instructional practices in reading and math. The results of this study will be compared with results from the Rasch analyses to investigate the congruence of the item difficulty ordering between the two groups of participants.

Study 4: Pilot of Instrument

Participants

Thirty elementary school teachers presently teaching in schools in the midwest participated in this study. All participants were female. Two participants were teaching kindergarten, six were teaching first grade, three were teaching second, six were teaching third, seven were teaching fourth, and six were teaching fifth grade.

Instrument and Procedures

Each participant was asked to complete the pilot instrument in Appendix F. This instrument is slightly different than the one in Appendix B due to changes made from the results of the content validity studies. The directions in part one were changed for clarity and a few of the items were revised. I asked the practicing teachers to complete the instrument and provided an additional sheet at the end of the instrument that asked them questions such as: (a) Were any of the directions unclear? If yes, which ones and do you have any suggestions? (b) Were any of the items unclear? If yes, which ones, why, and do you have any suggestions? (c) Was it easy to respond to each item using the scale provided? If not, why? and (d) Any other comments or suggestions?

The participants completed the form voluntarily and independently within the time period of one week. I ran preliminary item analysis procedures for each of the hypothesized constructs including means, standard deviations, ranges, and Cronbach alpha reliability coefficients. This allowed me to identify any items or scales that did not initially seem to be performing as expected so that changes could be made prior to preparation of the final instrument. Any comments received were evaluated and changes were made to the instrument where necessary or appropriate.

Results

For the item and subscale reliability analyses the items were divided by their operational categories (i. e., reading efficacy expectations, math efficacy expectations, reading outcome expectations, math outcome expectations). The means, standard deviations, ranges, subscale reliability, item-scale correlations, and alpha if item deleted for the reading efficacy expectation items (i. e., part one, items one through twenty four in Appendix F) are presented in Table 4. The scale reliability was .93, and the only item with a poor item-scale correlation was item twenty-three (integrating e-mail), but the alpha if item deleted does not indicate that the reliability would be greatly enhanced if this item were deleted. This is probably due to the large number of items for this scale.

The means, standard deviations, ranges, subscale reliability, item-scale correlations, and alpha if item deleted for the math efficacy expectation items (i. e., part one, items twenty-five through forty-eight in Appendix F) are provided in

Table 5. The scale reliability was .94, and two items (forty-four [author multimedia] and forty-eight [use World Wide Web]) had somewhat poor item-scale correlations, but the alpha if item deleted does not indicate that the reliability would be greatly enhanced if either of these items were deleted.

The means, standard deviations, ranges, subscale reliability, item-scale correlations, and alpha if item deleted for the reading outcome expectation items (i. e., part two, items one through twenty four in Appendix F) are included in Table 6. The scale reliability was .93, and none of the items had a poor item-scale correlation.

The means, standard deviations, ranges, subscale reliability, item-scale correlations, and alpha if item deleted for the math outcome expectation items (i. e., part two, items twenty-five through forty-eight in Appendix F) are presented in Table 7. The scale reliability was .91, and the only item with a poor item-scale correlation was item thirty (providing timed practice tests), but the alpha if item deleted does not indicate that the reliability would be greatly enhanced if this item were deleted. Because the item-scale correlation is negative this item may need to be reverse scored in the future.

Only a few participants provided comments or suggestions for improvement of the instrument. Some commented on the length of the instrument. They felt that it was too long and "repetitive." Two other comments pertained specifically to the response scale. One participant noted, "The 'degree of confidence' was rather confusing." Another participant said, "I didn't really see the point in filling out the last two pages because everyone should have confidence that all different ways of teaching can be useful. Some will work better for some people, but all can be used

Table 4

Means, standard deviations, ranges, item-scale correlations, and alpha if item deleted for reading efficacy expectation items

Standardized Cronbach Coefficient Alpha = .93

Item	M	SD	Range	Correlation with Total	Alpha if Item Deleted
1	4.57	.63	3 - 5	.63	.93
2	4.63	.67	3 - 5	.58	.93
3	4.40	.72	3 - 5	.72	.93
4	3.90	.99	2 - 5	.71	.93
5	4.37	.81	3 - 5	.76	.93
6	4.37	.76	2 - 5	.44	.93
7	4.70	.53	3 - 5	.48	.93
8	4.57	.82	2 - 5	.82	.93
9	4.63	.61	3 - 5	.80	.93
10	4.53	.86	2 - 5	.81	.93
11	4.47	.68	3 - 5	.82	.93
12	4.67	.55	3 - 5	.45	.93
13	4.73	.58	3 - 5	.75	.93
14	4.63	.56	3 - 5	.74	.93
15	4.33	.66	3 - 5	.50	.93
16	3.80	1.10	1 - 5	.74	.93
17	3.83	1.12	2 - 5	.74	.93
18	3.87	1.14	1 - 5	.59	.93
19	3.80	1.13	2 - 5	.35	.93
20	4.57	.63	3 - 5	.26	.94
21	2.83	1.37	1 - 5	.47	.93
22	3.73	1.31	1 - 5	.68	.93
23	3.63	1.43	1 - 5	.07	.94
24	3.30	1.49	1 - 5	.28	.93

Table 5

Means, standard deviations, ranges, item-scale correlations, and alpha if item deleted for math efficacy expectation items

Standardized Cronbach Coefficient Alpha = .94

Item	M	SD	Range	Correlation with Total	Alpha if Item Deleted
25	4.19	1.08	1 - 5	.76	.93
26	4.45	.89	2 - 5	.55	.94
27	4.26	.93	2 - 5	.70	.94
28	3.87	1.06	2 - 5	.72	.93
29	4.42	.85	2 - 5	.68	.94
30	4.84	.45	3 - 5	.58	.94
31	4.61	.67	3 - 5	.69	.94
32	4.65	.55	3 - 5	.51	.94
33	4.52	.77	2 - 5	.68	.94
34	4.35	1.05	1 - 5	.87	.93
35	4.58	.81	2 - 5	.71	.93
36	4.84	.45	3 - 5	.48	.94
37	4.65	.75	2 - 5	.78	.93
38	4.61	.72	3 - 5	.82	.93
39	3.97	1.02	1 - 5	.78	.93
40	3.87	1.14	1 - 5	.64	.94
41	4.03	1.02	1 - 5	.65	.94
42	4.10	1.02	1 - 5	.63	.94
43	4.35	.91	2 - 5	.61	.94
44	2.55	1.29	1 - 5	.20	.94
45	3.97	1.08	1 - 5	.53	.94
46	3.29	1.35	1 - 5	.39	.94
47	2.87	1.26	1 - 5	.46	.94
48	2.94	1.39	1 - 5	.14	.94

Table 6

Means, standard deviations, ranges, item-scale correlations, and alpha if item deleted for reading outcome expectation items

Standardized Cronbach Coefficient Alpha = .93

Item	M	SD	Range	Correlation with Total	Alpha if Item Deleted
1	4.84	.37	4 - 5	.66	.93
2	4.84	.45	3 - 5	.52	.93
3	4.69	.47	4 - 5	.62	.93
4	4.50	.80	2 - 5	.62	.93
5	4.66	.60	3 - 5	.71	.93
6	4.28	.99	2 - 5	.29	.93
7	4.84	.37	4 - 5	.64	.93
8	4.50	.72	3 - 5	.51	.93
9	4.84	.45	3 - 5	.68	.93
10	4.84	.45	3 - 5	.69	.93
11	4.69	.64	3 - 5	.73	.93
12	4.84	.37	4 - 5	.36	.93
13	4.81	.47	3 - 5	.38	.93
14	4.88	.34	4 - 5	.73	.93
15	4.47	.76	3 - 5	.74	.93
16	4.25	.84	2 - 5	.61	.93
17	4.25	.84	3 - 5	.85	.92
18	4.09	.93	2 - 5	.69	.93
19	3.91	1.28	1 - 5	.50	.93
20	4.38	.98	1 - 5	.48	.93
21	3.47	1.39	1 - 5	.55	.93
22	4.13	1.01	1 - 5	.41	.93
23	3.75	1.27	1 - 5	.39	.93
24	3.81	1.40	1 - 5	.50	.93

Table 7

Means, standard deviations, ranges, item-scale correlations, and alpha if item deleted for math outcome expectation items

Standardized Cronbach Coefficient Alpha = .91

Item	M	SD	Range	Correlation with Total	Alpha if Item Deleted
25	4.65	.55	3 - 5	.65	.90
26	4.77	.50	3 - 5	.47	.91
27	4.71	.53	3 - 5	.37	.91
28	4.42	.89	2 - 5	.46	.91
29	4.58	.62	3 - 5	.63	.90
30	4.06	.96	2 - 5	-.01	.92
31	4.52	.63	3 - 5	.60	.91
32	4.45	.81	2 - 5	.62	.90
33	4.77	.50	3 - 5	.43	.91
34	4.77	.50	3 - 5	.52	.91
35	4.68	.65	3 - 5	.78	.90
36	4.68	.54	3 - 5	.45	.91
37	4.87	.34	4 - 5	.28	.91
38	4.84	.45	3 - 5	.58	.91
39	4.23	.76	2 - 5	.71	.90
40	4.19	.91	2 - 5	.63	.90
41	4.23	.84	3 - 5	.82	.90
42	4.35	.66	3 - 5	.60	.91
43	4.45	.72	3 - 5	.67	.90
44	3.16	1.39	1 - 5	.43	.91
45	4.16	.97	1 - 5	.49	.91
46	3.94	1.18	1 - 5	.35	.91
47	3.74	1.12	1 - 5	.61	.91
48	3.61	1.50	1 - 5	.39	.91

and will work somewhere." These "last two pages" were those items assessing reading and math outcome expectations.

Discussion

The item analyses and subscale reliabilities calculated indicated high internal consistency and only a few items with poor item-scale correlations. Because of the small sample size used for item analyses even those items with poor item-scale correlations were not removed from the study. The content of these items seemed to fit with the intended sampling from the domain of instructional activities but most were identified as difficult in the previous content validity study. This may have contributed to the poor item-total correlation, and these items may not be identified as poor in the context of the purposes of Rasch analysis procedures. The high means, especially on those items not related to technology integration, and low standard deviations are possibly due to the high confidence that this sample of teachers possesses. Or, it may be that the percent confidence divisions are too wide or difficult for participants to differentiate. This concern was also proposed and addressed in study one.

While Bandura most often uses a percent confidence response scale such as the one I employed, all of the teacher efficacy instruments reviewed use a response scale of "strongly disagree" to "strongly agree." A revision of the pilot instrument might employ this response scale to investigate whether it produces greater variation in participant responses. Such a scale might also address the comments made by participants about the confusion over the "degree of confidence" response scale and its specific use with the outcome expectation items.

The length of the instrument is also cause for concern. The negative comments about the length and my desire for an adequate return rate on the final administration of the instrument influenced my decision to divide the instrument into two instruments, one measuring reading efficacy and outcome expectations and one measuring math efficacy and outcome expectations.

Studies Assessing Construct Validity, Criterion-related Validity and Reliability

For the construct and criterion-related validity, and reliability analyses, participants were a randomly drawn sample of 1500 elementary school teachers (K-6) in the state of Tennessee. Elementary school teachers were chosen due to the likelihood that they are responsible for teaching both math and reading to a single class of heterogeneously grouped students. The inclusion of middle or high school teachers would necessitate the construction of additional instruments assessing each of the classes and groups of students the teachers teach. Teachers in the state of Tennessee were chosen because of the statewide mandated testing of elementary students in math and reading and the availability of the database which provides an assessment of teachers' past and present achievements through measures of their students' gain scores in reading and math.

The design of this portion of the project used a mail out survey to encourage participation through anonymity. A consent form was not necessary since return of the survey in the envelope provided was indication of consent. An introduction letter explaining the purposes of the study to the teachers was included as a cover letter to the instrument (see Appendix G). A survey was mailed to each participant midsemester in Spring 1997. Seven hundred and fifty surveys each of reading and math were mailed. Follow-up postcards were mailed two weeks after the initial mailing to those who had not yet returned their survey. All survey data received within nine weeks of the initial mailing were included in the data analyses.

The final versions of the two instruments used (i. e., one for reading and one for math) are in Appendix H and I. These appendices provide the complete instruments used in the mail out survey including the demographic and criterion-related validity questions. Some additional questions were also asked for use in later analyses not reported in this project. Also note that I revised the response scale as proposed in the discussion of study four, and the efficacy and outcome expectation items were ordered to reflect the difficulty order obtained in study three.

The following studies report the participants, instruments and procedures used to analyze (a) construct validity and reliability using Rasch analyses for both the reading and the math instruments, (b) construct validity using confirmatory factor analyses for both instruments, and (c) criterion-related validity for both instruments.

Study 5: Rasch Analyses of Teacher Efficacy Instrument (Reading)

Participants

Of the 750 instruments mailed 268 were returned for a return rate of 36%. Of the 268 instruments, 209 were returned completed. The 59 that were returned incomplete were most often accompanied with hand written notes from the participants indicating that they did not teach reading in their current teaching assignment. The final pool of 209 participants included 200 females and 9 males with a mean age of 43 years (SD = 9.2, ranging from 24 - 63 years of age) and an average of 17 years of teaching experience (SD = 9.26, ranging from 2 - 38 years of teaching). The participants included 175 Caucasians, 17 African Americans, 1 American Indian, and 1 Asian American with 15 not reporting. Instruments were returned from a fairly even distribution of teachers at each grade level (K=32, 1=38, 2=30, 3=36, 4=19, 5=20, 6=14, with 20 not reporting). The average number of years participants had been teaching at their present grade level was 9 years (SD = 7.87, ranging from 1 - 35 years).

Instrument and Procedures

The Reading Efficacy and Outcome Expectations Instrument can be viewed in Appendix H. The 209 instruments completed and returned were analyzed using the computer program BIGSTEPS (Linacre & Wright, 1995). Analyses were performed for each set of twenty four items (i. e., one set hypothesized to measure reading efficacy expectations and one set hypothesized to measure reading outcome expectations) under the assumption that the data fit the Rasch Rating Scale Model. First, I examined the overall item fit indices to indicate overall fit of the data to the model. Second, I examined the individual item fit statistics to identify misfitting items for possible elimination. The criteria for misfit used throughout was mean square Infit and Outfit values outside the range of .6 - 1.4. Third, the previous two steps were repeated with misfitting items excluded until no items were identified as misfitting or they were within the approximated Type I error rate of 5% (i. e., one item in each set).

Once I obtained the final set of items, I examined (a) the average measures and step calibration measures for order to investigate the use of the response scale categories, (b) the item and person separation indices to investigate precision, and (c) the extent to which the items defined the underlying continuum from easy to difficult and if participants were sufficiently dispersed along the same continuum. The person ability measures for each construct were used in later criterion-related validity analyses.

Results

Reading Efficacy Expectation Items. The initial calibration of the twenty-four reading efficacy expectation items indicated the data fit the model well with global mean square Infit and Outfit statistics for items of .97 and 1.05, respectively, and for persons of 1.02 and 1.06, respectively. Examination of the individual item Infit and Outfit statistics (see Table 8) revealed three items (i. e., 4, 5, and 10) with values clearly outside of the acceptable range. These items are potentially measuring a construct other than that defined by the remaining items. These three items were dropped, and I performed a second calibration of the remaining twenty-one items.

Recalibration of the twenty-one reading efficacy expectation items indicated the data fit the model well with global mean square Infit and Outfit statistics for items of .96 and 1.01, respectively, and for persons of 1.00 and 1.02, respectively. Examination of the individual item Infit and Outfit statistics (see Table 9) revealed one item (item 3) with an outfit value (1.58) outside of the acceptable range. Since one item is within the approximated 5% Type I error rate, no items were eliminated and analysis of the reading efficacy expectations scale continued with these 21 items.

The average measures and step calibration measures were both in ascending order (see Table 9). The average measures are the average ability across all items of

Table 8

Initial Item Fit Statistics in Outfit Order for Reading Efficacy Expectations Scale

Item	Measure	Error	Infit Mean Square	Outfit Mean Square
4. integrating reading activities	-1.12	.12	.95	1.87
5. phonics skills	-.73	.11	1.51	1.71
10. choose literature books	.02	.08	1.48	1.49
3. whole group instruction	-.99	.11	1.30	1.45
14. learning experiences for diverse learners	.45	.07	.90	1.33
20. literature books on CD-ROM	.64	.07	1.19	1.20
13. challenging seatwork	-.28	.09	.95	1.22
23. use World Wide Web	1.57	.06	1.12	1.12
24. author multimedia	1.84	.06	1.06	1.11
21. use e-mail	1.44	.06	1.09	1.06
7. drill and practice software	.30	.08	1.00	1.06
12. word processing activities	-.50	.10	1.11	1.02
1. cooperative groups	-.31	.09	.92	1.01
9. small group instruction	-.97	.11	1.07	.99
2. integrating writing activities	-.97	.11	.90	.90
6. plan effective lessons	-1.13	.12	.98	.84
19. adjust to learning needs	-.34	.09	.74	.81
18. adjust to student skills	-.37	.09	.73	.78
22. organize activities according to student background	.43	.08	.79	.80
11. prepare engaging materials	-.68	.10	.77	.69
8. instruction based on student interest	-.41	.10	.68	.69
15. selecting computer software	.83	.07	.75	.73
16. use computer tutorials	.73	.07	.71	.72
17. activities using computers	.57	.07	.69	.66
MEAN	.00	.09	.97	1.05
S.D.	.86	.02	.23	.32

Table 9

Final Item Fit Statistics in Difficulty Order for Reading Efficacy Expectations Scale

Item	Measure	Error	Infit Mean Square	Outfit Mean Square
24. author multimedia	1.85	.07	1.03	1.09
23. use World Wide Web	1.57	.07	1.08	1.07
21. use e-mail	1.43	.07	1.06	1.03
15. selecting computer software	.79	.07	.74	.72
16. use computer tutorials	.68	.07	.70	.70
20. literature books on CD-ROM	.59	.07	1.22	1.24
17. activities using computers	.52	.08	.68	.65
14. learning experiences for diverse learners	.38	.08	.96	1.40
22. organize activities according to student background	.36	.08	.85	.88
7. drill and practice software	.23	.08	1.02	1.09
13. challenging seatwork	-.39	.10	1.01	1.31
1. cooperative groups	-.42	.10	.99	1.11
19. adjust to learning needs	-.45	.10	.79	.85
18. adjust to student skills	-.49	.10	.80	.85
8. instruction based on student interest	-.53	.10	.76	.80
12. word processing activities	-.62	.10	1.18	1.13
11. prepare engaging materials	-.82	.11	.85	.76
2. integrating writing activities	-1.11	.11	.96	.98
9. small group instruction	-1.12	.12	1.14	1.10
3. whole group instruction	-1.14	.12	1.40	1.58
6. plan effective lessons	-1.29	.12	1.04	.90
All Items				
MEAN	.00	.09	.96	1.01
S.D.	.91	.02	.18	.24
All Persons				
MEAN	1.31	.29	1.00	1.02
S.D.	1.07	.11	.72	.83

	Category 1	2	3	4	5	6
Average Measures	-1.02	-.55	-.09	.61	1.36	2.52
Step Calibration Measures		-.85	-.80	-.59	.19	2.05

persons responding in that response category. Ascending order is a desirable characteristic because you would expect that, on average, people with lower ability (i. e., efficacy or outcome expectations) would respond to items using a lower response category. Order of step measures indicates that each response category was the most probable response at some point along the response continuum. Participants used each response category in stepwise fashion and did not find it easy to skip a lower response category in favor of a higher category.

The scale had an item separation and reliability of 9.60 and .99 and person separation and reliability of 2.92 and .91. The scale identified 13.13 distinct item difficulty strata that the participants distinguished and 4.23 distinct reading efficacy strata distinguished by the items. Both of these are above the criteria of two strata needed for the scale to be useful in distinguishing individuals with high efficacy and low efficacy.

Table 9 provides the items in difficulty order from most difficult to least difficult along with their measures and other statistics. Figure 2 presents the map of items against persons. Persons with higher reading efficacy and items more difficult to endorse are located toward the top of the map with persons of lower reading efficacy and items easier to endorse near the bottom. Both of these were used to assess the extent to which the items defined the underlying continuum from easy to difficult and if participants were sufficiently dispersed along the same continuum. The range of item calibrations was from 1.85 (item 24) to -1.29 (item 6) logits. Investigation of the ordering of these items provided evidence of construct validity given that the items seemed to be logically ordered from least to most difficult and were relatively congruent with results from content validity study three. Notable

```
SCALE
MEASURE        PERSONS      MAP OF ITEMS (DIFFICULTY ORDER)

   6               ##  +                          M = Mean
                       |                          S = 1 S.D.
                    .  |                          Q = 2 S.D.
   5                   +                          # ≈ 3 Persons
                       |                          . = 1 Person
                   .#  |
                       |
                    .  |
   4                .  +
                       |
                 # Qp  |
                   .#  |
   3                #  +
                   ##  |
               .# Sp   |
                 .##   |
   2              ###  +
            .#####     |QI   author multimedia
            .#####     |     use e-mail; use World Wide Web
       .######## Mp    |
   1  .#########       +SI
          #######      |     computer tutorials; selecting software
           .#####      |     diverse learner; using computers; CD-ROMs
           .#### Sp    |     drill and practice software; student background
   0       .####       +MI
                #      |     learning needs; cooperative groups; seatwork
                .      |     word processing; student interest; student skills
             . Qp      |     prepare engaging materials
  -1            .      +SI   whole group; small group; integrating writing
                       |     plan effective lessons
                       |
                       |QI
  -2                   +
```

Figure 2. Map of Items and Persons for Reading Efficacy Expectations Scale

exceptions were items 1 (cooperative groups) and 7 (drill and practice software) that were identified within the moderately difficult range rather than the easy range, and adjusting to learning needs (item 19) and student skills (item 18) being ranked as less difficult than they were in study three.

Conceptually, I could group the items into three difficulty groups. The easiest items to agree with were those assessing efficacy for the more global, generic, or common instructional activities used by most teachers, including whole and small group instruction, planning effective lessons, and preparing engaging materials. The common use of these activities is likely to lead teachers to more easily agree that they have confidence in their ability to employ these strategies. The more difficult items to agree with were those assessing efficacy for instructional activities related to individualizing instruction, including organizing activities according to student background or based on student interests, adjusting activities to individual learning needs, and adjusting activities to account for individual differences in skill. These activities certainly require more knowledge and skill as well as effort and time, and it is therefore possible that teachers would find it more difficult to agree that they have confidence in their ability to employ these strategies when teaching reading. The most difficult items for participants to endorse were those instructional activities related to integrating technology into reading instruction, including using e-mail and literature on CD-ROM, selecting computer software, and providing activities for use of the World Wide Web or multimedia authoring tools. These activities require specific training, knowledge, and skills related to technology which many teachers may not possess or may be in the early stages of developing considering the recent spread of and encouragement for technology use in classrooms. Thus, these are the least likely items to which teachers are willing to agree that they have confidence in their ability to employ these instructional activities. The two exceptions to this grouping schema were technology-related items 12 (word processing) and 7 (drill and practice software) which were identified as easy and moderately difficult, respectively. These are the most common uses of computers in schools and may have resulted in teachers ranking them as less difficult.

Although several item difficulty strata and person efficacy strata were distinguished, the items do not span the entire continuum of person ability. The average item measure was 0.0 logits (SD = .91) and the average person measure was 1.31 logits (SD = 1.07). Only a few of the most difficult items (i. e., items 21, 23, and 24) are above the average person measure indicating that the scale is failing to assess the higher end of the reading efficacy continuum.

Reading Outcome Expectation Items. The initial calibration of the twenty-four reading outcome expectation items indicated the data fit the model well with global mean square Infit and Outfit statistics for items of .97 and 1.03, respectively, and for persons of 1.03 and 1.02, respectively. Examination of the individual item Infit and Outfit statistics (see Table 10) revealed four items (i. e., 3, 5, 9, and 13) with values clearly outside of the acceptable range. These items are potentially measuring a construct other than that defined by the remaining items. Item 17 also had values just outside the lower end of the acceptable range of .6. However, Wright & Linacre (1994) suggested leaving items outside the lower end of the range in upon recalibration because often these items will recalibrate within the acceptable range when other items above the high end of the range are eliminated. Therefore

Table 10

Initial Item Fit Statistics in Outfit Order for Reading Outcome Expectations Scale

Item	Measure	Error	Infit Mean Square	Outfit Mean Square
13. challenging seatwork	.30	.09	1.73	2.23
5. phonics skills	-.45	.10	1.76	1.82
3. whole group instruction	.09	.09	1.43	1.49
9. small group instruction	-.23	.10	1.38	1.47
7. drill and practice software	.50	.09	.97	1.10
24. author multimedia	1.16	.08	.93	1.06
14. learning experiences for diverse learners	-.32	.10	1.10	1.06
1. cooperative groups	-.07	.10	1.00	1.04
10. choose literature books	.24	.09	.99	1.00
2. integrating writing activities	-.37	.10	.70	.97
19. adjust to learning needs	-.85	.11	.87	.96
16. use computer tutorials	.35	.09	.84	.96
20. literature books on CD-ROM	.32	.09	.90	.95
11. prepare engaging materials	-.79	.11	.91	.91
18. adjust to student skills	-.67	.11	.88	.91
22. organize activities according to student background	.55	.09	.88	.91
12. word processing activities	-.18	.10	.97	.85
21. use e-mail	1.41	.07	.80	.84
8. instruction based on student interest	-.18	.10	.82	.81
23. use World Wide Web	1.20	.08	.74	.75
6. plan effective lessons	-1.35	.13	.73	.72
4. integrating reading activities	-1.37	.13	.74	.67
15. selecting computer software	.35	.09	.63	.64
17. activities using computers	.37	.09	.59	.58
MEAN	.00	.10	.97	1.03
S.D.	.71	.01	.30	.37

Table 11

Final Item Fit Statistics in Difficulty Order for Reading Outcome Expectations Scale

Item	Measure	Error	Infit Mean Square	Outfit Mean Square
21. use e-mail	1.63	.08	.82	.86
23. use World Wide Web	1.38	.08	.74	.74
24. author multimedia	1.33	.08	.96	1.08
22. organize activities according to student background	.61	.09	1.08	1.08
7. drill and practice software	.55	.09	1.15	1.40
17. activities using computers	.41	.10	.61	.59
15. selecting computer software	.38	.10	.66	.67
16. use computer tutorials	.38	.10	.90	1.00
20. literature books on CD-ROM	.34	.10	.93	.94
10. choose literature books	.26	.10	1.17	1.20
1. cooperative groups	-.10	.10	1.20	1.54
8. instruction based on student interest	-.23	.11	.97	.93
12. word processing activities	-.23	.11	1.02	.86
14. learning experiences for diverse learners	-.39	.11	1.25	1.18
2. integrating writing activities	-.45	.11	.82	1.04
18. adjust to student skills	-.79	.12	1.03	1.08
11. prepare engaging materials	-.93	.12	1.09	1.11
19. adjust to learning needs	-1.00	.12	1.02	1.24
6. plan effective lessons	-1.56	.13	.85	1.01
4. integrating reading activities	-1.58	.13	.85	.78
All Items				
MEAN	.00	.10	.96	1.02
S.D.	.88	.01	.17	.23
All Persons				
MEAN	1.88	.35	.99	1.01
S.D.	1.34	.14	.70	.72

	Category 1	2	3	4	5	6
Average Measures	-.97	-.65	.01	.87	1.77	3.40
Step Calibration Measures		-1.32	-1.36	-.77	.66	2.79

only the four items outside the upper range of 1.4 were dropped, and I performed a second calibration of the remaining twenty items.

Recalibration of the twenty reading outcome expectation items indicated the data fit the model well with global mean square Infit and Outfit statistics for items of .96 and 1.02, respectively, and for persons of .99 and 1.01, respectively. Examination of the individual item Infit and Outfit statistics (see Table 11) revealed that item 17 fell within the acceptable range upon recalibration. However, one item (item 1) had its Outfit value (1.54) outside of the acceptable range. Since one item is within the approximated 5% Type I error rate, no items were eliminated and analysis of the reading outcome expectations scale continued with these 20 items.

The average measures were in ascending order indicating that persons responding to each successive response category had increasing average ability measures. However, the step calibration measure between response category 2 and response category 3 was out of order (see Table 11) indicating that response category 2 was never the most probable response at some point along the response continuum. Participants found it easier to respond to the items with "somewhat disagree" rather than "disagree" after deciding not to respond with "strongly disagree." This lack of ordered response categories may indicate the need to combine or omit categories from the response scale.

The scale had an item separation and reliability of 8.16 and .99 and person separation and reliability of 3.14 and .91. The scale identified 11.21 distinct item difficulty strata that the participants distinguished and 4.52 distinct reading outcome strata distinguished by the items. Both of these are above the criteria of two strata needed for the scale to be useful in distinguishing individuals with high outcome and low outcome expectations.

Table 11 provides the items in difficulty order from most difficult to least difficult along with their measures and other statistics. Figure 3 presents the map of items against persons. Persons with higher reading outcome expectations and items more difficult to endorse are located toward the top of the map with persons of lower reading outcome expectations and items easier to endorse near the bottom. Both of these were used to assess the extent to which the items defined the underlying continuum from easy to difficult and if participants were sufficiently dispersed along the same continuum. The range of item calibrations was from 1.63 (item 21) to -1.58 (item 4) logits. Investigation of the ordering of these items provided evidence of construct validity given that the items seemed to be logically ordered from least to most difficult to endorse.

Once again the items could conceptually be grouped using the global, individualized, and technology groupings of the reading efficacy expectation items. However, an interesting difference should be noted. There was some overlap of the global activities with the activities related to individualizing instruction on the easy to endorse end of the continuum. This overlap seems logical when considering that the outcome expectations are assessing teachers' confidence that these activities will positively influence their students' reading achievement. It makes sense that teachers would easily agree that they were confident that individualizing their instruction for each student according to their skills or learning needs would positively influence reading achievement. Exceptions to these groupings were item 12 (word processing) which was identified as moderately difficult and item 22 (student background) which was identified as difficult to

```
SCALE
MEASURE    PERSONS        MAP OF ITEMS (DIFFICULTY ORDER)

   7          .##  +                          M = Mean
                   |                          S  = 1 S.D.
                   |                          Q  = 2 S.D.
   6            #  +                          #  ≈ 3 Persons
                   |                          .  = 1 Person
                #  |
   5           .#  +
               Qp  |
                #  |
   4           .#  +
              .##  |
            ## Sp  |
   3           ##  +
           .####   |
           .####   |
   2     .###### Mp+
          ######   |QI    use e-mail
    .###########   |      author multimedia; use World Wide Web
   1   #########   +SI    student background
        .#### Sp   |      using computers; drill & practice software
              .##  |      choose lit. books; CD-ROM; comp. tutor.; select software
   0          .##  +MI    cooperative groups
                .  |      integ. writing; div. learners; word proc.; stud. interest
            .# Qp  |      student skills
  -1               +SI    learning needs; prepare engaging materials
                   |
                .  |QI    integrating reading; plan effective lessons
  -2               +
                   |
  -3               +
                   |
  -4            .  +
```

Figure 3. Map of Items and Persons for Reading Outcome Expectations Scale

endorse. Apparently, teachers do not feel that organizing activities around student backgrounds is as likely to influence achievement as other individualization strategies such as adjusting to learning needs and student skills.

Again, although several item difficulty strata and person efficacy strata were distinguished, the items do not span the entire continuum of person ability. The average item measure was 0.0 logits (SD = .88) and the average person measure was 1.88 logits (SD = 1.34). Even the most difficult item to endorse (item 21 with a measure of 1.63 logits) did not exceed the average person measure indicating that the scale is failing to assess the higher end of the reading outcome continuum.

Discussion

The above analyses provide strong evidence for construct validity and reliability for reading efficacy and reading outcome expectations according to the item fit, item difficulty hierarchy, and separation indices. While the content of the reading efficacy items that were dropped (i. e., phonics skills, choosing own literature books, and integrating Reading) were initially thought to be part of the same construct as the other items, the item fit statistics indicated otherwise. The high Infit and Outfit values indicated excessive variability on these items. This is produced by a surprising number of teachers scoring higher than expected on the item, and a surprising number of others scoring lower than expected. This indicates that the items are not contributing consistently to defining reading efficacy or outcome expectations measured by the other items. The way these items are measuring teachers' efficacy and outcome expectations for teaching reading is not the same as the way reading efficacy and outcome expectations are marked out by the remaining items on each scale which fit together to provide general definitions of the constructs. Why these items do not fit statistically could be the result of any number of possible causes, but these data do not indicate which is at work here.

One possible explanation for these three items misfitting may be that teachers are either very likely or very unlikely to engage in these activities depending on the instructional method they use to teach reading. If a teacher is using a basal reader approach to teaching reading, he or she may be less likely to have students choose their own literature books for reading instruction or purposefully integrate reading activities into other curriculum areas. But if a teacher is using a whole language or integrated curriculum approach to teaching reading, he or she may be more likely to integrate Reading activities across the curriculum and have students choose their own literature books, and less likely to provide phonics skills instruction. These differences in strategies based on the instructional method employed may contribute to excessive variance in teachers' responses to these items.

The instructional method teachers use and the activities they are likely to employ are often not determined by themselves, but by administrators or school boards or current thinking and theory related to teaching reading. So, it may be difficult to make the leap from saying that the methods teachers use determine their level of confidence in employing these strategies. However, Bandura's theory proposes that past experience is the most influential source of efficacy. If teachers have employed specific instructional strategies in the course of using a specific method for teaching reading, and they have experienced success with these strategies, then they are likely to have higher confidence in their ability to employ these strategies.

A similar reasoning might be used to account for the misfitting items on the reading outcome expectations scale. The content of the reading outcome items that were dropped (i. e., phonics skills, whole group instruction, small group instruction, and challenging seatwork) were initially thought to be part of the same construct as the other items, but the item fit statistics indicated otherwise. The high Infit and Outfit values indicated excessive variability on these items. Once again, depending on the instructional method used to teach reading, teachers may believe one or more of these strategies to be very influential or not influential at all in improving student achievement in reading.

The slight variation of the instructional activities related to individualizing along the difficulty continuum are important to note. These teachers found it easier to endorse that individualized instruction will positively influence student reading achievement, but it was more difficult for them to agree that they had confidence in their ability to employ these strategies. These lower reading efficacy expectations for individualizing instruction are likely to be the better predictor of the effort and persistence teachers may exert to employ such strategies. However, their willingness to admit that such activities are likely to positively influence achievement (i. e., higher outcome expectations for individualized instruction items) may indicate that teachers might be highly motivated to gain additional knowledge and training on how to implement strategies for individualizing instruction.

Both sets of items determining the reading efficacy and reading outcome scales did not seem to adequately assess the higher end of the difficulty continuum. This may indicate that additional items need to be developed that will assess the higher end of the efficacy and outcome continuums. The possibility also exists that these teachers, and possibly teachers in general, are likely to report that they are very confident in their ability to employ a variety of instructional activities when teaching reading and are also confident that most strategies will positively influence reading achievement. This tendency for teachers to report overconfidence in teaching has been previously documented (Lanier & Little, 1986; Zimmerman, 1996).

Another concern is the response scale for the reading outcome expectations scale. The lack of step order indicates participants did not use category 2 (disagree) as expected by the model. Further analysis to find the optimal categorization for this scale may be a focus for future research.

Study 6: Rasch Analyses of Teacher Efficacy Instrument (Mathematics)

Participants

Of the 750 instruments mailed 257 were returned for a return rate of 34%. Of the 257 instruments, 180 were returned completed. The 77 that were returned incomplete were most often accompanied with hand written notes from the participants indicating that they did not teach math in their current teaching assignment. The final pool of 180 participants included 167 females and 13 males with a mean age of 43 years (SD = 10.07, ranging from 23 - 61 years of age) and an average of 16 years of teaching experience (SD = 9.20, ranging from 1 - 40 years of teaching). The participants included 163 Caucasians and 12 African Americans with 5 not reporting. Instruments were returned from a fairly even distribution of teachers at each grade level (K=26, 1=27, 2=33, 3=26, 4=20, 5=21, 6=15, with 12 not reporting). The average number of years participants had been teaching at their present grade level was 10 years (SD = 8.10, ranging from 1 - 38 years).

Instrument and Procedures

The Mathematics Efficacy and Outcome Expectations Instrument can be viewed in Appendix I. The 180 instruments completed and returned were analyzed using the same computer program and same procedures described above for the Reading Instrument.

Results

Mathematics Efficacy Expectation Items. The initial calibration of the twenty-four math efficacy expectation items indicated the data fit the model well with global mean square Infit and Outfit statistics for items of .96 and .99, respectively, and for persons of 1.07 and .99, respectively. Examination of the individual item Infit and Outfit statistics (see Table 12) revealed two items (i. e., 5 and 21) with both values outside of the acceptable range. These items are potentially measuring a construct other than that defined by the remaining items. These two items were dropped, and I performed a second calibration of the remaining twenty-two items.

Recalibration of the twenty-two math efficacy expectation items indicated the data fit the model well with global mean square Infit and Outfit statistics for items of .93 and .97, respectively, and for persons of 1.05 and .97, respectively. Examination of the individual item Infit and Outfit statistics (see Table 13) revealed one item (item 24) with both values outside of the acceptable range. Since one item is within the approximate 5% Type I error rate, no items were eliminated and analysis of the mathematics efficacy expectations scale continued with these 22 items.

The average measures were in ascending order indicating that persons responding to each successive response category had increasing average ability

Table 12

Initial Item Fit Statistics in Outfit Order for Mathematics Efficacy Expectations Scale

Item	Measure	Error	Infit Mean Square	Outfit Mean Square
21. use spreadsheet activities	1.38	.07	1.41	1.62
5. timed practice tests	-.66	.12	1.64	1.61
24. author multimedia	2.10	.07	1.35	1.44
22. use World Wide Web	1.84	.07	1.14	1.18
6. drill and practice software	.59	.08	1.25	1.18
18. learning experiences for diverse learners	.55	.08	.87	1.15
4. hands-on manipulatives	-1.00	.13	1.13	1.13
9. challenging seatwork	-.68	.12	.97	1.08
17. selecting computer software	.74	.08	1.00	1.04
20. organize activities according to student background	.37	.09	.87	1.03
3. cooperative groups	-.56	.11	1.09	1.01
12. activities using computers	.63	.08	1.06	.98
23. use videodiscs	1.08	.07	.91	.95
16. use computer tutorials	.66	.08	.82	.84
14. use computer games	.48	.08	.88	.83
11. prepare engaging materials	-.70	.12	.78	.81
8. plan effective lessons	-.44	.11	.86	.81
7. integrate math activities	-1.17	.14	.64	.77
2. whole group instruction	-1.72	.16	.94	.76
1. provide worked examples	-1.54	.15	.81	.75
15. adjust to student skills	-.12	.10	.69	.75
19. instruction based on student interest	.01	.09	.60	.75
10. small group instruction	-1.47	.15	.74	.63
13. adjust to learning needs	-.35	.11	.59	.62
MEAN	.00	.10	.96	.99
S.D.	1.02	.03	.26	.27

Table 13

Final Item Fit Statistics in Difficulty Order for Mathematics Efficacy Expectations Scale

Item	Measure	Error	Infit Mean Square	Outfit Mean Square
24. author multimedia	2.21	.07	1.47	1.56
22. use World Wide Web	1.94	.07	1.18	1.21
23. use videodiscs	1.15	.07	.97	1.06
17. selecting computer software	.80	.08	.98	.99
16. use computer tutorials	.71	.08	.82	.87
12. activities using computers	.68	.08	1.06	.98
6. drill and practice software	.64	.08	1.25	1.18
18. learning experiences for diverse learners	.60	.08	.90	1.15
14. use computer games	.53	.08	.88	.83
20. organize activities according to student background	.42	.09	.90	1.10
19. instruction based on student interest	.05	.10	.61	.79
15. adjust to student skills	-.09	.10	.72	.78
13. adjust to learning needs	-.33	.11	.62	.66
8. plan effective lessons	-.42	.11	.88	.87
3. cooperative groups	-.55	.11	1.15	1.13
9. challenging seatwork	-.67	.12	1.00	1.11
11. prepare engaging materials	-.69	.12	.78	.81
4. hands-on manipulatives	-1.00	.13	1.15	1.16
7. integrate math activities	-1.18	.14	.64	.77
10. small group instruction	-1.49	.15	.75	.62
1. provide worked examples	-1.56	.15	.84	.82
2. whole group instruction	-1.74	.16	.97	.81
All Items				
MEAN	.00	.10	.93	.97
S.D.	1.05	.03	.21	.22
All Persons				
MEAN	1.64	.30	1.05	.97
S.D.	1.07	.11	.69	.51

	Category 1	2	3	4	5	6
Average Measures	-.92	-.56	-.03	.64	1.50	2.78
Step Calibration Measures		-.98	-.48	-.67	.23	1.90

measures. But, the step calibration measure between response category 3 and response category 4 was out of order (see Table 13) indicating that response category 3 was never the most probable response at some point along the response continuum. Participants found it easier to respond to the items with "somewhat agree" rather than "somewhat disagree" after deciding not to respond with "disagree." This lack of ordered response categories may indicate the need to combine or omit categories from the response scale.

The scale had an item separation and reliability of 9.49 and .99 and person separation and reliability of 2.77 and .88. The scale identified 12.99 distinct item difficulty strata that the participants distinguished and 4.03 distinct math efficacy strata distinguished by the items. Both of these are above the criteria of two strata needed for the scale to be useful in distinguishing individuals with high efficacy and low efficacy.

Table 13 provides the items in difficulty order from most to least difficult along with their measures and other statistics. Figure 4 presents the map of items against persons. Persons with higher mathematics efficacy and items more difficult to endorse are located toward the top of the map with persons of lower mathematics efficacy and items easier to endorse near the bottom. Both of these were used to assess the extent to which the items defined the underlying continuum from easy to difficult and if participants were sufficiently dispersed along the same continuum. The range of item calibrations was from 2.21 (item 24) to -1.74 (item 2) logits. Investigation of the ordering of these items provided evidence of construct validity given that the items seemed to be logically ordered from least to most difficult and were relatively congruent with results from content validity study three. A notable

```
SCALE
MEASURE       PERSONS        MAP OF ITEMS (DIFFICULTY ORDER)

   6              ##   +                            M = Mean
                       |                            S  = 1 S.D.
                   .   |                            Q  = 2 S.D.
   5                   +                            #  ≈ 3 Persons
                   .   |                            .  = 1 Person
                   #   |
   4              ##   +
                 .# Qp |
                 .##   |
                  .#   |
   3             ###   +
            .#### Sp   |
                  .#   |
              .######  |    author multimedia
   2            ####   +QI  use World Wide Web
         .######## Mp  |
         #########     |
        .#########     |    use videodiscs
   1     #########     +SI
        .#########     |    drill & prac. soft.; use comp.; comp.tut.; sel. software
            .##### Sp  |    stud. background; computer games; div. learners
             #####     |
   0               #   +MI  student skills; student interest
                  .#   |    learning needs
                 . Qp  |    cooperative groups; plan effective lessons
                       |    prepare engaging materials; seatwork
  -1                   +SI  hands-on manipulatives
                       |    integrate math
                       |    worked examples; small group
                       |    whole group
  -2                   +QI
```

Figure 4. Map of Items and Persons for Mathematics Efficacy Expectations Scale

exception is that item 6 (drill and practice software) was identified as difficult rather than easy as in the content validity study.

Conceptually, I could group the items into three difficulty groups. Just as with the reading efficacy items, the easiest items to agree with were those assessing efficacy for the more global, generic, or common instructional activities used by most teachers, including whole and small group instruction, using hands-on manipulatives, planning effective lessons, and providing worked examples. The common use of these activities is likely to lead teachers to more easily agree that they have confidence in their ability to employ these strategies. The more difficult items to agree with were those assessing efficacy for instructional activities related to individualizing instruction, including organizing activities according to student background or based on student interests, adjusting activities to individual learning needs, and providing learning experiences for diverse learners. These activities certainly require more knowledge and skill as well as effort and time, and it is therefore possible that teachers would find it more difficult to agree that they have confidence in their ability to employ these strategies when teaching math. The most difficult items for participants to endorse were those instructional activities related to integrating technology into math instruction, including using videodiscs and computer games, selecting computer software, and providing activities for use of the World Wide Web or multimedia authoring tools. These activities require specific training, knowledge, and skills related to technology which many teachers may not possess or may be in the early stages of developing considering the recent spread of and encouragement for technology use in classrooms. Thus, these are the least likely items to which teachers are willing to agree that they have confidence in their ability to employ these instructional activities.

Although several item difficulty strata and person efficacy strata were distinguished, the items do not span the entire continuum of person ability. The average item measure was 0.0 logits (SD = 1.05) and the average person measure was 1.64 logits (SD = 1.07). Only a few of the most difficult items (i. e., items 22 and 24) are above the average person measure indicating that the scale is failing to assess the higher end of the mathematics efficacy continuum.

Mathematics Outcome Expectation Items. The initial calibration of the twenty-four mathematics outcome expectation items indicated the data fit the model well with global mean square Infit and Outfit statistics for items of .96 and 1.01, respectively, and for persons of 1.02 and 1.00, respectively. Examination of the individual item Infit and Outfit statistics (see Table 14) revealed four items (i. e., 5, 9, 19 and 20) with values outside of the acceptable range. These items are potentially measuring a construct other than that defined by the remaining items. These four items were dropped, and I performed a second calibration of the remaining twenty items.

Recalibration of the twenty mathematics outcome expectation items indicated the data fit the model well with global mean square Infit and Outfit statistics for items of .95 and 1.03, respectively, and for persons of 1.01 and 1.03, respectively. Examination of the individual item Infit and Outfit statistics (see Table 15) revealed that one item (item 4) had an Outfit value (1.64) outside the acceptable range upon recalibration. Since one item is within the hypothesized 5% Type I error rate, no

Table 14

Initial Item Fit Statistics in Outfit Order for Mathematics Outcome Expectations Scale

Item	Measure	Error	Infit Mean Square	Outfit Mean Square
9. challenging seatwork	.45	.10	1.75	1.80
5. timed practice tests	.24	.10	1.52	1.62
19. instruction based on student interest	.41	.10	1.40	1.50
20. organize activities according to student background	.59	.09	1.33	1.48
4. hands-on manipulatives	-.75	.12	1.16	1.27
22. use World Wide Web	1.25	.08	1.06	1.18
24. author multimedia	1.34	.08	.98	1.10
18. learning experiences for diverse learners	-.35	.11	1.02	1.06
8. plan effective lessons	-.39	.11	.94	1.02
1. provide worked examples	-.79	.12	1.03	1.00
10. small group instruction	-.83	.12	.87	1.00
3. cooperative groups	-.08	.10	.95	1.00
2. whole group instruction	-.44	.11	.93	.94
13. adjust to learning needs	-.89	.12	.86	.89
6. drill and practice software	.28	.10	.86	.85
11. prepare engaging materials	-.83	.12	.65	.81
23. use videodiscs	.89	.09	.80	.79
15. adjust to student skills	-.53	.11	.74	.78
14. use computer games	.00	.10	.77	.77
21. use spreadsheet activities	1.38	.08	.71	.75
17. selecting computer software	.16	.10	.75	.72
16. use computer tutorials	.24	.10	.70	.69
12. activities using computers	.13	.10	.65	.64
7. integrate math activities	-1.47	.14	.69	.62
MEAN	.00	.10	.96	1.01
S.D.	.74	.01	.28	.31

Table 15

Final Item Fit Statistics in Difficulty Order for Mathematics Outcome Expectations Scale

Item	Measure	Error	Infit Mean Square	Outfit Mean Square
21. use spreadsheet activities	1.70	.09	.82	.88
24. author multimedia	1.65	.09	1.09	1.19
22. use World Wide Web	1.55	.09	1.12	1.22
23. use videodiscs	1.11	.10	.86	.85
6. drill and practice software	.41	.11	.99	.97
16. use computer tutorials	.35	.11	.76	.76
17. selecting computer software	.27	.11	.79	.78
12. activities using computers	.23	.11	.64	.63
14. use computer games	.08	.11	.80	.79
3. cooperative groups	-.01	.11	1.05	1.07
18. learning experiences for diverse learners	-.31	.12	1.13	1.24
8. plan effective lessons	-.36	.12	1.13	1.34
2. whole group instruction	-.41	.12	1.11	1.16
15. adjust to student skills	-.51	.12	.82	.89
4. hands-on manipulatives	-.76	.12	1.28	1.64
1. provide worked examples	-.81	.12	1.22	1.25
10. small group instruction	-.86	.13	.97	1.26
11. prepare engaging materials	-.86	.13	.68	1.03
13. adjust to learning needs	-.92	.13	.96	1.03
7. integrate math activities	-1.56	.14	.73	.68
All Items				
MEAN	.00	.11	.95	1.03
S.D.	.90	.01	.18	.25
All Persons				
MEAN	2.13	.35	1.01	1.03
S.D.	1.18	.09	.67	.76

	Category 1	Category 2	Category 3	Category 4	Category 5	Category 6
Average Measures	-1.36	-.33	.17	1.11	2.11	3.40
Step Calibration Measures		-2.08	-1.10	-.71	.98	2.92

items were eliminated and analysis of the mathematics outcome expectations scale continued with these 20 items.

The average measures and step calibration measures were both in ascending order (see Table 15) indicating that persons responding to each successive response category had increasing average ability measures and each response category was the most probable response at some point along the response continuum. Participants used each response category in stepwise fashion and did not find it easy to skip a lower response category in favor of a higher category.

The scale had an item separation and reliability of 7.63 and .98 and person separation and reliability of 2.79 and .89. The scale identified 10.50 distinct item difficulty strata that the participants distinguished and 4.05 distinct mathematics outcome strata distinguished by the items. Both of these are above the criteria of two strata needed for the scale to be useful in distinguishing individuals with high outcome and low outcome expectations in math.

Table 15 provides the items in difficulty order from most difficult to least difficult along with their measures and other statistics. Figure 5 presents the map of items against persons. Persons with higher mathematics outcome expectations and items more difficult to endorse are located toward the top of the map with persons of lower mathematics outcome expectations and items easier to endorse near the bottom. Both of these were used to assess the extent to which the items defined the underlying continuum from easy to difficult and if participants were sufficiently dispersed along the same continuum. The range of item calibrations was from 1.65 (item 24) to -1.56 (item 7) logits. Investigation of the ordering of these items provided evidence of construct validity given that the items seemed to be logically

```
SCALE
MEASURE        PERSONS        MAP OF ITEMS (DIFFICULTY ORDER)

  7              ###  +                              M = Mean
                      |                              S  = 1 S.D.
                   .  |                              Q  = 2 S.D.
  6                   +                              #  ≈ 3 Persons
                      |                              .  = 1 Person
                      |
  5                #  +
                  .#  |
                  # Qp|
  4             .###  +
              .#####  |
               ### Sp|
  3          .######  +
             #######  |
             #######  |
  2     ############ Mp+
       .############  |QI  use WWW; author multimedia; spreadsheet act.
        ############  |
  1            ##### Sp+SI  use videodiscs
                 ###  |
                  ##  |    use comp.; select soft.; comp.tut.; drill & prac. soft.
  0                #  +MI  cooperative groups; computer games
                 ## Qp|    whole group; effective lessons; diverse learners
                      |    worked examples; hands-on manip.; student skills
 -1                   +SI  learning needs; engaging materials; small group
                   .  |
                      |QI  integrate math
 -2                   +
```

Figure 5. Map of Items and Persons for Mathematics Outcome Expectations Scale

ordered from least to most difficult to endorse.

Once again the items could conceptually be grouped using the global, individualized, and technology groupings of the mathematics efficacy expectation items. However, the same difference noted between the reading efficacy and reading outcome items was also found in these analyses. There was some overlap of the global activities with the activities related to individualizing instruction on the easy to endorse end of the continuum. This overlap seems logical when considering that the outcome expectations are assessing teachers' confidence that these activities will positively influence their students' mathematics achievement. It makes sense that teachers would easily agree that they were confident that individualizing their instruction for each student according to their skills or learning needs would positively influence mathematics achievement.

Again, although several item difficulty strata and person efficacy strata were distinguished, the items do not span the entire continuum of person ability. The average item measure was 0.0 logits (SD = .90) and the average person measure was 2.13 logits (SD = 1.18). Even the most difficult item to endorse (item 24 with a measure of 1.65 logits) did not exceed the average person measure indicating that the scale is failing to assess the higher end of the mathematics outcome expectations continuum.

Discussion

The above analyses provide strong evidence for construct validity and reliability for mathematics efficacy and mathematics outcome expectations according to the item fit, item difficulty hierarchy, and separation indices. While the content of the mathematics efficacy items that were dropped (i. e., timed practice tests and spreadsheet activities) were initially thought to be part of the same construct as the other items, the item fit statistics indicated otherwise. The high Infit and Outfit values indicated excessive variability on these items. This is produced by a surprising number of teachers scoring higher than expected on the item, and a surprising number of others scoring lower than expected. This indicates that the items are not contributing consistently to defining mathematics efficacy or outcome expectations measured by the other items. The way these items are measuring teachers' efficacy and outcome expectations for teaching mathematics is not the same as the way mathematics efficacy and outcome expectations are marked out by the remaining items on each scale which fit together to provide general definitions of the constructs. Why these items do not fit statistically could be the result of any number of possible causes, but these data do not indicate which is at work here.

While an explanation for misfitting items is only an educated guess, one possible explanation for the misfit of the "timed practice tests" item may be that teachers are either very likely or very unlikely to engage in this activity depending on their views of the appropriateness of the strategy. In the past timed practice tests were often used as a method of sharpening students' computational skills and preparing them for timed standardized achievement tests. Presently, current thinking in student motivation discourages use of such strategies due to their focus on performance rather than on learning. These two views that teachers possibly possess may result in excessive variance in responses to such an item, whether it is assessing an efficacy or outcome expectation. A final note on the "timed practice tests" item that was dropped in both scales is that this item was also identified in the item analyses of the pilot instrument as having a poor item-scale correlation.

Obviously, this item seems to be assessing a construct other than the one intended. The misfit of the "spreadsheet activities" could be due to the poor clarity of the item content. My intention was to focus on the use of computerized spreadsheet programs in mathematics instruction. Review of the item content indicated that this focus was probably not as clear as it should have been.

The content of the mathematics outcome items that were dropped (i. e., timed practice tests, instruction based on student interest, instruction based on student background, and challenging seatwork) were initially thought to be part of the same construct as the other items, but the item fit statistics indicated otherwise. The high Infit and Outfit values indicated excessive variability on these items. Interestingly, the "challenging seatwork" item was also dropped in the reading outcome scale. A possible explanation is that teachers do not even consider that an activity such as seatwork has any bearing on student achievement in either math or reading. Or, it might be these teachers interpreted "challenging" to mean "hard" and they do not think that hard seatwork helps students learn. In contrast to the reading outcome scale, teachers were not likely to include individualized instruction according to student interest or background as appropriate for influencing achievement in the mathematics domain. It may be that teachers feel students' interest and background have more bearing on their understanding of what they are reading than on their understanding of mathematical concepts. Or, it may be that teachers could not think of strategies they use that specifically incorporate student interests or backgrounds, so they found it difficult to link these strategies to student achievement in math.

Once again, the slight variation of the instructional activities related to individualizing along the difficulty continuum are important to note. These teachers found it easier to endorse that individualized instruction will positively influence student mathematics achievement, but it was more difficult for them to agree that they had confidence in their ability to employ these strategies. These lower mathematics efficacy expectations for individualizing instruction are likely to be the better predictor of the effort and persistence teachers may exert to employ such strategies. However, their willingness to admit that such activities are likely to positively influence achievement (i. e., higher outcome expectations for individualized instruction) may indicate that teachers could be highly motivated to gain additional knowledge and training on how to implement strategies for individualizing instruction in math.

As with the reading scales, both sets of items determining the math efficacy and math outcome scales did not seem to adequately assess the higher end of the difficulty continuum. This may indicate that additional items need to be developed that will assess the higher end of the efficacy and outcome continuums. The possibility also exists that these teachers, and possibly teachers in general, are likely to report that they are very confident in their ability to employ a variety of instructional activities when teaching math and are also confident that most strategies will positively influence math achievement.

A final concern is the response scale for the mathematics efficacy expectations scale. The lack of step order indicates participants did not use category 3 (somewhat disagree) as expected by the model. Further analysis to find the optimal categorization for this scale may be a focus for future research.

Study 7: Confirmatory Factor Analyses of Teacher Efficacy Instruments (Reading and Mathematics)

The purpose of factor analysis is to examine the interrelationships among instrument items and verify clusters of items that share sufficient variation to justify their existence as a construct being measured by the instrument. Confirmatory factor analysis (CFA) allows specification of a unique factorial solution, statistically evaluates the fit of the model to the data, and suggests possible modifications for model improvement (Mueller, 1996). Hambleton and Swaminathan (1985) noted, however, that the use of correlations could lead to a factor solution with too many factors, and that some of them may be "difficulty factors" found because of the range of item difficulties among the items in the test. Similar concerns are voiced by Banerji, Smith, and Dedrick (1997) who caution researchers to carefully consider the analytic tools to use when examining properties of a scale and make a choice based upon the purposes of the scaling. When these points are considered it seems possible that CFA may not yield a desirable solution due to the identification of difficulty factors.

Participants, Instruments and Procedures

Participants and the instruments used for this study were described previously in studies five and six and will not be repeated here. For the present study, two confirmatory factor analyses (CFA) were conducted. The first CFA used the forty-one items on the Reading Efficacy and Outcome Expectations Instrument remaining from the Rasch analyses. The second CFA used the forty-two items on the Mathematics Efficacy and Outcome Expectations Instrument remaining from the Rasch analyses. I used the computer program EQS (Bentler, 1995) with correlation matrices as input and maximum likelihood as the estimation procedure. Each CFA specified that (a) responses could be explained by a two factor model; (b) the two factors were correlated; (c) each item would be free to load on one specified factor, with zero loadings on the other factor; and (d) the error terms for the items would be uncorrelated.

Results

Reading Efficacy and Outcome Expectations Instrument. The hypothesized two-factor model for the reading instrument did not provide satisfactory fit to the data [χ^2 (778) = 4625.01, $p < .001$; CFI = .52; GFI = .36; AGFI = .29]. The reported chi-square is significant indicating a poor fit. The ratio of chi-square to df is approximately 6:1 which is greater than the 5:1 rule of thumb for acceptable model-data fit. In addition, all other indices are well below the .90 needed for indication of acceptable fit. Given this inadequate model-data fit, I decided to examine the multivariate Lagrange Multiplier (LM) tests, which identify fixed parameters that if estimated would lead to improved model fit. These tests indicated specific error terms that if allowed to covary would improve model-data fit.

The examination of the multivariate LM tests revealed that correlating error terms of items related to one another by their difficulty rank established in the Rasch analyses would improve model fit. Table 16 provides the first 24 sets of error terms identified by the multivariate LM tests to improve model-data fit and illustrates this hypothesized conclusion. Those pairs marked by an "*" were the difficult items among the reading efficacy items. Those pairs marked by a "#" were those easier efficacy items near the lower end of the difficulty continuum provided in Table 9 (see page 94). A similar pattern holds for the reading outcome items.

Table 16

Multivariate Lagrange Multiplier Tests for Reading Instrument

Step	Parameter	Instrument Item	Chi-square	D.F.	Probability
1	E19,E18	REE19,REE18	147.143	1	0.0001
$ 2	E47,E45	ROE23,ROE21	287.000	2	0.0001
+ 3	E43,E42	ROE19,ROE18	422.100	3	0.0001
+ 4	E42,E38	ROE18,ROE14	553.847	4	0.0001
* 5	E23,E21	REE23,REE21	671.866	5	0.0001
* 6	E24,E23	REE24,REE23	795.750	6	0.0001
+ 7	E43,E38	ROE19,ROE14	906.657	7	0.0001
$ 8	E48,E47	ROE24,ROE23	1013.481	8	0.0001
$ 9	E48,E45	ROE24,ROE21	1123.533	9	0.0001
* 10	E24,E21	REE24,REE21	1223.448	10	0.0001
+ 11	E35,E30	ROE11,ROE6	1323.316	11	0.0001
+ 12	E32,E26	ROE8,ROE2	1417.763	12	0.0001
+ 13	E43,E26	ROE19,ROE2	1508.135	13	0.0001
# 14	E11,E6	REE11,REE6	1594.334	14	0.0001
+ 15	E42,E26	ROE18,ROE2	1679.471	15	0.0001
+ 16	E42,E32	ROE18,ROE8	1766.123	16	0.0001
+ 17	E43,E32	ROE19,ROE8	1850.099	17	0.0001
+ 18	E30,E28	ROE6,ROE4	1930.977	18	0.0001
# 19	E11,E8	REE11,REE8	2010.221	19	0.0001
# 20	E12,E11	REE12,REE11	2089.137	20	0.0001
# 21	E11,E9	REE11,REE9	2165.708	21	0.0001
# 22	E12,E2	REE12,REE2	2235.118	22	0.0001
+ 23	E38,E32	ROE14,ROE8	2303.466	23	0.0001
+ 24	E38,E26	ROE14,ROE2	2372.083	24	0.0001

Note. REE=Reading Efficacy Expectation Item, ROE=Reading Outcome Expectation Item, "*" = difficult reading efficacy items, "#" = easy reading efficacy items, "$" = difficult reading outcome items, "+" = easy reading outcome items.

Those pairs marked by a "$" were the difficult items on the reading outcome scale (see Table 11, page 100). Those pairs marked by a "+" indicate that the error terms of those easier reading outcome items should be allowed to correlate to improve model-data fit. This pattern continued throughout the rest of the multivariate LM tests with errors grouping together according to their difficulty level. Apparently, the results of the CFA are unsatisfactory because the covariation of items due to their difficulty level are indicating that additional factors (perhaps "difficulty" factors) should be included in the model. This finding leads me to assume that CFA may be an inappropriate technique for assessing dimensionality when a continuum of difficulty is underlying that dimension. Therefore, no further model modifications were made and the poor model-data fit was accepted as an idiosyncrasy of the constructs I was attempting to assess. My hypothesis regarding the poor model-data fit due to the continuum of difficulty underlying the reading efficacy and outcome dimensions warrants further investigation in later studies.

Mathematics Efficacy and Expectations Instrument. The hypothesized two-factor model for the math instrument also did not provide satisfactory fit to the data [χ^2 (818) = 3426.64, $p < .001$; CFI = .52; GFI = .39; AGFI = .32]. The reported chi-square is significant indicating a poor fit. However, the ratio of chi-square to df is approximately 4:1 which is less than the liberal ratio of 5:1 rule of thumb for satisfactory model-data fit. But, all other indices are well below the .90 needed for indication of acceptable fit. Examination of the multivariate LM tests revealed much of the same pattern as these tests for the reading instrument. Correlating error terms of items related to one another by their difficulty rank established in the Rasch analyses would improve model-data fit. Table 17 provides

Table 17

Multivariate Lagrange Multiplier Tests for Reading Instrument

Step	Parameter	Instrument Item	Chi-square	D.F.	Probability
+ 1	E39,E37	MOE15,MOE13	99.490	1	0.0001
# 2	E2,E1	MEE2,MEE1	195.034	2	0.0001
+ 3	E35,E31	MOE11,MOE7	282.225	3	0.0001
4	E24,E22	MEE24,MEE22	367.355	4	0.0001
# 5	E11,E7	MEE11,MEE7	435.512	5	0.0001
* 6	E19,E15	MEE19,MEE15	500.507	6	0.0001
* 7	E20,E19	MEE20,MEE19	566.527	7	0.0001
$ 8	E42,E38	MOE18,MOE14	621.161	8	0.0001
+ 9	E34,E26	MOE10,MOE2	675.104	9	0.0001
$ 10	E42,E36	MOE18,MOE12	724.961	10	0.0001
# 11	E11,E10	MEE11,MEE10	773.897	11	0.0001
12	E48,E46	MOE24,MOE22	822.421	12	0.0001
* 13	E20,E15	MEE20,MEE15	870.821	13	0.0001
+ 14	E26,E25	MOE2,MOE1	918.865	14	0.0001
$ 15	E27,E26	MOE3,MOE2	967.607	15	0.0001
* 16	E19,E13	MEE19,MEE13	1015.327	16	0.0001
* 17	E15,E13	MEE15,MEE13	1061.534	17	0.0001
* 18	E15,E14	MEE15,MEE14	1108.723	18	0.0001
+ 19	E35,E34	MOE11,MOE10	1153.831	19	0.0001
+ 20	E34,E31	MOE10,MOE7	1200.527	20	0.0001
+ 21	E31,E26	MOE7,MOE2	1246.912	21	0.0001
# 22	E7,E2	MEE7,MEE2	1290.604	22	0.0001
# 23	E10,E7	MEE10,MEE7	1335.094	23	0.0001
# 24	E8,E7	MEE8,MEE7	1379.039	24	0.0001

Note. MEE=Mathematics Efficacy Expectation Item, MOE=Mathematics Outcome Expectation Item, "*" = moderately difficult math efficacy items, "#" = easy math efficacy items, "$" = moderately difficult math outcome items, "+" = easy math outcome items.

the first 24 sets of error terms identified by the multivariate LM tests to improve model-data fit and illustrates this hypothesized conclusion. Those pairs marked by an "*" were the moderately difficult of the math efficacy items falling in the middle of the continuum in Table 13 (see page 110). Those pairs marked by a "#" were those easy efficacy items near the lower end of the difficulty continuum provided in Table 13. A similar pattern holds for the math outcome items. Those pairs marked by a "$" were the moderately difficult items on the math outcome scale (see Table 15, page 116). Those pairs marked by a "+" indicate that the error terms of those easy math outcome items should be allowed to correlate to improve model-data fit. This pattern continued throughout the rest of the multivariate LM tests. Again, the results of the CFA are unsatisfactory because the covariation of items due to their difficulty level are indicating that additional factors (possibly "difficulty" factors) should be included in the model. No further model modifications were made and the poor model fit was accepted as an idiosyncrasy of the constructs I was attempting to assess.

Discussion

Evaluation of the construct validity for these instruments using confirmatory factor analyses was inappropriate when considering the theoretical basis of scale construction used to develop these instruments. Banerji, Smith, and Dedrick (1997) suggested that "the applicability of factor analytic or Rasch techniques for examining dimensionality should be decided based upon the purposes of scaling and the processes used to operationalize the construct" (p. 81). The purpose of scale construction in this project was to be able to scale items along a continuum of easy to endorse to difficult to endorse. When operationalizing the constructs of teacher efficacy and outcome expectations, I developed the content of the items with the goal of intentionally creating a continuum of easy to difficult instructional strategies or activities and I attempted to order the items by difficulty on the hypothesized scales.

Properties of item order like I intended are not well detected by correlational techniques such as factor analysis (Banerji, Smith, & Dedrick, 1997). In fact, factor analysis procedures often group items based on their difficulty rather than shared content (Hambleton & Swaminathan, 1985). As evidenced by the multivariate LM tests, the CFAs identified groups of items that should be allowed to correlate in order to improve model-data fit. These items were grouped according to their relative level of difficulty as identified by the Rasch procedures. Logically, participants' responses to these items falling at similar levels of difficulty would likely correlate with one another. However, in such instances where difficulty is not the dimension of concern but the continuum along which items were ordered, CFA is unable to detect the dimensionality of the construct of interest. Rasch models are much more appropriate and useful when the process of scale construction was deliberately designed to yield an ordering of items that fits the specifications of the Rasch model (Banerji, Smith, & Dedrick, 1997).

The poor results obtained from the above CFA procedures provided clear evidence for the need to heed Banerji, Smith, and Dedrick's (1997) suggestions. Selection of analytic tools for examining properties of a scale should be done carefully to avoid misleading results. In particular, researchers in the area of efficacy should be careful when examining the dimensionality of their instruments, especially if they have attempted to tap Bandura's efficacy dimension of magnitude by deliberately including items of varying difficulty levels.

Study 8: Criterion-related Validity

Participants, Instruments and Procedures

Participants for this study were described previously in studies five and six and will not be repeated here. Measures of efficacy and outcome expectations are theorized to correlate with behavioral reports or measures of effort, persistence, and choice. Therefore, several additional items were added to the survey instrument and administered simultaneously in order to assess these behaviors. These items are included with each instrument in Appendices H and I. Each instrument included items related to effort, persistence, and choice in each content domain and also related to technology integration. The technology-related items were assessed in the event that these efficacy and outcome expectation items were identified as tapping a different dimension. The previous analyses did not indicate the need to separate the technology-related items and eliminated the need for using the technology-related criterion items. Also, a portion of the teachers failed to complete these items due to no access to technology or in the event that another teacher was responsible for this area of instruction. Therefore, these criterion-related validity items related to technology were not used in the following analyses. Those that were used are included in Table 18 for clarity.

The answer for the first item indicated teachers' preference for teaching the four major content areas. Based on theory I hypothesized that a teacher high in efficacy expectations for a specific content area would be more likely to rank that content area as the preferred one; therefore I expected a negative correlation between the

Table 18

Criterion-related Validity Items for Assessing Effort, Persistence, and Task Preference

Item and Response Format	SD	D	SWD	SWA	A	SA
1. Please rank order your preference for teaching the following subjects to your students. The subject you **most prefer** to teach should receive a rank of "1". Your next subject preference should receive a rank of "2" and so on up to a rank of "4". Science______ Math_____ Reading______ Social Studies ______						
2. Students often vary in the way they participate and engage in activities in their class. In your opinion, approximately what **percentage** of the students in your class are actively engaged in the **reading (math)** lessons on a typical day? ________%						
3. Approximately how much time (in minutes) do you spend in a typical week **planning and preparing materials** for teaching Reading (Math) ? ______						
4. Approximately how much time (in minutes) do you spend in a typical week **engaged with your students in activities** related to Reading (Math)?_____						
5. Compared to the normal effort I exert teaching Reading (Math), I put more effort into working with a student who is not making progress in his/her reading ability.	1	2	3	4	5	6
6. Compared to the normal effort I exert teaching Reading (Math), I put more effort into developing enrichment activities for students who have above average reading ability.	1	2	3	4	5	6
7. Compared to the normal effort I exert planning in other subject areas, I put more effort into planning daily Reading (Math) instruction.	1	2	3	4	5	6
8. I continue working with a student until he or she learns essential reading (math) skills.	1	2	3	4	5	6
9. I continue to try a variety of instructional strategies until I find a strategy that is effective with my students in Reading (Math).	1	2	3	4	5	6

(table continues)

Item and Response Format						
10. I try a variety of approaches to teaching Reading (Math) in order to discover strategies that motivate my students to improve their reading ability.	1	2	3	4	5	6
11. I feel I have control over my students' achievement in Reading (Math).	1	2	3	4	5	6
12. Teaching Reading (Math) is stressful.	1	2	3	4	5	6
13. I think teachers in my building can affect student achievement in Reading (Math) more than me.	1	2	3	4	5	6
14. I feel responsible for my students' achievement in Reading (Math).	1	2	3	4	5	6

efficacy expectation measures and the preference rank for the corresponding content area. Teachers' perceptions of student engagement during instruction in a specific content area (item 2) was hypothesized to be positively correlated with the corresponding efficacy and outcome expectation measures. The third and fourth items asked teachers to report the amount of time (in minutes) they spend planning for and engaged in instruction in reading or math during a week. Bandura's theory and researchers in teacher efficacy have suggested that teachers who have greater efficacy and outcome expectations in a given area will spend more time planning for and engaging in instruction in that content area. Thus, positive correlations were expected between each of the measures of teacher planning and teacher engagement and their corresponding efficacy and outcome expectations measures.

To investigate the relationships between measures of reading and math efficacy and outcome expectations and teachers' effort and persistence in teaching reading or math, I developed three items each to assess effort and persistence in teaching reading or math (items five through seven and eight through ten, respectively, in Table 18). Theory predicts that each subscale measure of effort and persistence would have a positive correlation with its corresponding efficacy and outcome expectations measure. Additional items found in Table 18 assessed teachers' feelings of responsibility and control over their students' achievement, their beliefs about their effectiveness compared to other teachers, and the degree to which they felt teaching reading or math was stressful. These items were included due to previous research by Ashton (1984) and Ashton, Webb, and Doda (1983) who reported that these feelings or beliefs could be distinguished among high and low efficacy teachers and hypothesized that they would also be related to teacher effort and persistence. The stress and teacher comparison items were reverse scored prior to analyses. I hypothesized that each of these items would have a positive correlation with the corresponding efficacy and outcome measures. The participants responded to items five through fourteen using the same 6-point Likert scale as the efficacy and outcome expectation items. All correlations used the person ability measures for math and reading efficacy and outcome expectations obtained from the Rasch analyses.

Results

Since the low number of items measuring effort and persistence made it unlikely that Rasch analyses would yield meaningful results, subscale reliabilities were calculated for each of the three items measuring effort and persistence. This procedure resulted in very low subscale reliabilities between .29 and .38 for the four subscales in math and reading. This was partially due to the low number of items for each scale. Examination of the correlation matrices for each set of 10 criterion items revealed that the majority of correlations among the items were between .20 and .67. Because measures of effort and persistence are likely to be positively correlated and the correlations among many items were satisfactory, all items (items five through fourteen in Table 18) were combined and I performed Rasch analyses for each content area. The Rasch analyses assessed the unidimensionality of the items, allowed diagnosis of misfitting items, and provided an estimate of reliability. Rasch procedures also yielded person ability measures that could be used in the correlation analyses. These interval measures, rather than ordinal measures obtained from averaging the Likert responses, are more appropriate to use in statistical analyses.

The initial calibration identified the item assessing stress had poor fit for both reading and math. Apparently, stress is not tapping the same construct as the other effort and persistence-related items. The recalibration of each set of items with the stress item dropped for reading and math yielded satisfactory item fit and scale reliability for each. The final item and scale fit statistics for reading and math are presented in Tables 19 and 20.

Recalibration of the reading items indicated the data fit the model well with global mean square Infit and Outfit statistics for items of .99 and .99, respectively, and for persons of 1.03 and .99, respectively. Examination of the individual item Infit and Outfit statistics (see Table 19) revealed that item 5 (try strategies for effectiveness) had an Outfit value (.59) slightly outside of the acceptable range. Since one item is within the approximated 5% Type I error rate, no items were eliminated. The average measures were in ascending order indicating that persons responding to each successive response category had increasing average effort measures. However, the step calibration measure between response category 3 and response category 4 was out of order indicating that response category 3 was never the most probable response at some point along the response continuum. Participants found it easier to respond to the items with "somewhat agree" rather than "somewhat disagree" after deciding not to respond with "disagree." This lack of ordered response categories may indicate the need to combine or omit categories from the response scale. The scale had an item separation and reliability of 6.03 and .97 and person separation and reliability of 1.57 and .71. The scale identified 8.37 distinct item difficulty strata that the participants distinguished and 2.43 distinct reading effort strata distinguished by the items.

Table 19

Final Item Fit Statistics for Reading Effort Scale

Item	Measure	Error	Infit Mean Square	Outfit Mean Square
effort for planning instruction	.36	.08	1.20	1.29
effort for stud. w/poor progress	-.05	.08	1.24	1.29
effort for enrichment activities	.84	.07	1.15	1.25
comparison w/other teachers	.31	.08	1.23	1.20
responsible for achievement	-.36	.09	1.15	1.09
control over achievement	.59	.07	.83	.84
continue working w/student	-.16	.09	.68	.69
try approaches to motivate	-.90	.10	.76	.68
try strategies for effectiveness	-.62	.10	.63	.59
All Items				
MEAN	.00	.08	.99	.99
S.D.	.54	.01	.24	.27
All Persons				
MEAN	1.16	.42	1.03	.99
S.D.	.79	.08	.77	.69

Table 20

Final Item Fit Statistics for Mathematics Effort Scale

Item	Measure	Error	Infit Mean Square	Outfit Mean Square
comparison w/other teachers	.34	.08	1.25	1.27
effort for stud. w/poor progress	-.25	.09	1.25	1.22
effort for planning instruction	1.08	.08	1.13	1.13
responsible for achievement	-.94	.11	1.20	1.13
effort for enrichment activities	1.07	.08	.90	.88
control over achievement	.62	.08	.83	.85
try strategies for effectiveness	-.80	.11	.84	.83
try approaches to motivate	-.82	.11	.82	.79
continue working w/student	-.29	.09	.75	.75
All Items				
MEAN	.00	.09	1.00	.98
S.D.	.76	.01	.20	.19
All Persons				
MEAN	1.40	.41	1.00	.98
S.D.	.72	.07	.65	.65

Recalibration of the math items indicated the data fit the model well with global mean square Infit and Outfit statistics for items of 1.00 and .98, respectively, and for persons of 1.00 and .98, respectively. Examination of the individual item Infit and Outfit statistics (see Table 20) revealed no misfitting items. The average measures were not in ascending order and indicated that persons responding in category 1 (strongly disagree) had a higher average ability than persons responding in category 2 (disagree). However, the step calibration measures were in ascending order. The scale had an item separation and reliability of 7.72 and .98 and person separation and reliability of 1.51 and .69. The scale identified 10.63 distinct item difficulty strata that the participants distinguished and 2.35 distinct math effort strata distinguished by the items. The person ability (effort) measures obtained from each of these analyses were used in the following correlational analyses.

The means, standard deviations, and ranges were calculated for the efficacy and outcome expectations scales, the effort scales, student engagement, teacher planning, and teacher engagement for reading and mathematics. These are reported in Table 21. Overall, participants completing the reading instrument reported spending more time planning for and engaged in reading instruction than did the participants completing the math instrument. Participants completing the math instrument reported higher efficacy expectations, outcome expectations, and effort for teaching than participants completing the reading instrument.

Tables 22 and 23 report the correlations calculated between each of the criterion variables (subject preference rank, student engagement, teacher planning, teacher engagement, and teacher effort) and the efficacy and outcome expectations in each content domain, as well as the correlation between efficacy and outcome

Table 21

Descriptive Statistics for Criterion-related Validity Measures

Measure	Mean	SD	Range
Reading			
Efficacy Expectations	1.44	1.31	-1.07 - 6.18
Outcome Expectations	2.01	1.67	-5.42 - 6.80
Effort for Teaching Reading	1.16	.79	-1.60 - 3.54
Teacher Planning Time	111.96 min.	86.24	15.00 - 600.00
Teacher Engagement Time	389.45 min.	264.06	15.00 - 1000.00
Student Engagement Percentage	86.82 %	14.78	20.00 - 100.00
Mathematics			
Efficacy Expectations	1.74	1.26	-.42 - 6.20
Outcome Expectations	2.28	1.42	-1.30 - 7.00
Effort for Teaching Math	1.40	.72	-.22 - 5.38
Teacher Planning Time	96.92 min.	83.88	15.00 - 600.00
Teacher Engagement Time	270.02 min.	183.85	30.00 - 1000.00
Student Engagement Percentage	84.19 %	18.84	20.00 - 100.00

Table 22

Criterion-related Validity Correlational Analysis for Reading

	Effort	Preference Rank	Tchr Plan	Tchr Engage	Student Engage	Outcome
Efficacy	.29**	-.03	.03	.01	.16*	.54**
Outcome	.29**	.01	-.04	.10	.11	

* $p < .025$ ** $p < .0001$

Table 23

Criterion-related Validity Correlational Analysis for Mathematics

	Effort	Preference Rank	Tchr Plan	Tchr Engage	Student Engage	Outcome
Efficacy	.21*	-.04	-.03	.14	.10	.44**
Outcome	.46**	-.04	.01	.01	.02	

* $p < .005$ ** $p < .0001$

expectations. I obtained partial confirmation for the validity of the inferences from these instruments given the positive and significant correlations between the measures of effort and the measures of efficacy and outcome expectations.

Reading efficacy and reading outcome expectations were significantly and equally correlated with effort ($r = .29$ for both). However, only reading efficacy expectations were correlated with teachers' perceptions of the percentage of students engaged in reading lessons on a regular basis. Those teachers who perceived a higher percentage of their students were engaged during instruction also reported confidence in their ability to teach reading. There were no other significant correlations among the reading measures except for the expected positive correlation between reading efficacy and reading outcome expectations ($r = .54$).

Mathematics efficacy and mathematics outcome expectations were each significantly correlated with effort ($r = .21$ and $r =.46$, respectively). Here, mathematics outcome expectations are correlated higher with the measure of effort. This indicates that in the content domain of mathematics teachers are more likely to put forth effort when teaching if they have higher confidence in the ability of the instructional strategies to influence math achievement. There was also a positive correlation between the math efficacy and math outcome measures ($r = .44$).

Discussion

These results are encouraging in that they provide partial evidence for the validity of inferences made from these instruments. The positive and significant correlations of the efficacy and outcome expectation measures with measures of teachers' self-reports of effort in teaching allow us to hypothesize with some confidence that efficacy and outcome expectations as measured by these instruments can be used to predict the amount of effort and persistence teachers are likely to put forth when teaching. This evidence supports Bandura's expectancy theory and his hypothesis that efficacy and outcome expectations are related to reports of effort and persistence. Within this study, it is interesting to note that within the reading domain, efficacy and outcome expectations were equally correlated with measures of effort. However, within the mathematics domain, outcome expectations were more highly correlated with effort than efficacy expectations. The above findings do not support Bandura's hypothesis that measures of efficacy and outcome expectations are likely to influence choice or preference for tasks. While in all but one case the expected negative direction of the relationship between preference rank and efficacy or outcome expectations was found, the correlations were essentially zero. It may be that a more elaborate measure of choice or preference for teaching reading or math needs to be developed before we can make conclusions about its relationship to efficacy and outcome expectations. One item scales often have poor reliability resulting in artificially low correlations.

These results also do not support the validity of inferences from measures of efficacy and outcome expectations to the amount of time teachers spend planning for and engaged with their students in instruction in reading or mathematics. An observational analysis made during data entry may explain these nonsignificant results. Several participants made comments throughout the survey. An evaluation of these comments with particular attention to the time spent on planning revealed that a portion of teachers ($n = 28$ for the reading instrument and $n = 21$ for the math instrument) noted that they spent little time on planning due to their many years of experience. These comments might explain why poor correlations were obtained between planning time and efficacy and outcome expectations.

While the correlations were nonsignificant between teacher engagement and measures of efficacy and outcome expectations in reading or math, the pattern of correlations is of interest. On the reading instrument, measures of outcome expectations were more highly correlated ($r = .10$) with teacher engagement than efficacy expectations ($r = .01$). Potentially, teachers who have confidence in the strategies they use to teach reading (rather than confidence in their ability to teach reading) are more likely to spend more time engaged in reading instruction. The opposite pattern was found on the math instrument. Measures of mathematics efficacy expectations were more highly correlated ($r = .14$) with teacher engagement than outcome expectations ($r = .01$). In the mathematics domain, teachers who have confidence in their ability to teach math (rather than confidence in the strategies used) are more likely to spend more time engaged in math instruction with their students. A more elaborate or well-defined measure of time engaged might further clarify the relationships among these variables. One item scales often have poor reliability resulting in artificially low correlations. Also, comments from participants indicated that some teachers, specifically in the domain of reading, had a difficult time reporting the time they spent engaged in reading instruction because they "did reading all the time" or "integrated reading into all other subjects."

The correlations among efficacy and outcome expectation measures and teachers' perceptions of student engagement in instruction revealed one significant correlation between reading efficacy and student engagement rate ($r = .16$). The same pattern of correlations, although nonsignificant, was found between reading outcome and student engagement ($r = .11$), and math efficacy and student engagement ($r = .10$). However, teachers' confidence in the strategies used to teach math was not related to their perceptions of students' engagement in math instruction ($r = .02$).

As expected, there were strong positive correlations between efficacy and outcome expectations within each domain ($r = .54$ for reading and $r = .44$ for mathematics). These correlations are in strong contrast to the near zero or negative correlations found in previous measures of teacher efficacy and outcome expectations (Ashton, Webb, & Doda, 1983; Gibson & Dembo, 1984; Woolfolk & Hoy, 1990) but are in line with Bandura's theory (Bandura, 1986) and other research findings (Maddux, Norton, & Stoltenberg, 1986; Manning & Wright, 1983). This is easily explained through the different approach I used in measuring efficacy and outcome expectations. My approach was more aligned with Bandura and clearly distinguishable from other teacher efficacy and outcome measures.

While some have concluded that high correlations between efficacy and outcome measures make it unnecessary to measure outcome expectancies due to redundancy (Bandura, 1986; Manning & Wright, 1983), the different patterns of correlations found in this study indicate that each measure is distinct in its correlations with some of the criterion measures. Additionally, when reading effort measures are regressed on reading efficacy and reading outcome measures each make unique and significant contributions to prediction of effort ($\underline{R}^2 = .14$ [Adj. $\underline{R}^2 = .13$], with reading outcome, $\beta = .223$, $\underline{t} = 2.88$, $p < .004$, and reading efficacy, $\beta = .196$, $\underline{t} = 2.53$, $\underline{p} < .012$). However, when math effort measures are regressed on math efficacy and math outcome measures only math outcome expectations make a unique contribution to the prediction of effort ($R^2 = .22$ [Adj. $R^2 = .21$], with math outcome, $\beta = .461$, $t = 6.20$, $p < .0001$, and math efficacy, $\beta = .009$, $t = .120$, $p < .90$). These results, specifically within the reading domain, make the distinction and measurement of both expectancies necessary and provide further evidence of the need to measure outcome expectations, especially in the mathematics domain.

CHAPTER V

Conclusions

The goal of this project was to investigate the validity and reliability of inferences from instruments measuring teacher efficacy expectations and teacher outcome expectations in the content domains of reading and math. The studies and findings discussed in the previous chapter provide support for the use of these instruments in future research related to these constructs. The process of collecting evidence of reliability and validity is continual. This chapter provides a synthesis of the findings from the validation studies and then focuses on the direction of future research on these instruments and future use of these instruments that will provide further evidence of reliability and validity.

Synthesis of Findings

The goal of this project was to develop and examine the validity and reliability of inferences made from instruments measuring teacher efficacy expectations and teacher outcome expectations in the content domains of reading and mathematics. My intention was to remedy early empirical research flaws in the measurement of teacher efficacy expectations and teacher outcome expectations through the construction of a measure more clearly in alignment with Bandura's theory. Through the content, construct, and criterion-related validity analyses, as well as reliability analyses, the goals of this project were met.

All three of the content validity studies provided solid evidence that (a) the conceptual and operational definitions of teacher efficacy and teacher outcome expectations were aligned with Bandura's expectancy theory, (b) the items could be identified as measuring distinct categories or constructs, and (c) the items represented a range of difficulty. The Rasch analyses also provided further evidence of content validity through the congruence with the findings of the third content validity of the level of difficulty for the majority of items.

Rasch Rating Scale analyses were used to identify items that were misfitting and provided evidence for construct validity through the final item statistics and item difficulty hierarchy for each construct. The Rasch results also indicated areas in which the instruments need improvement, including further definition of items assessing the higher end of the efficacy and outcome continuums in both reading and math and further examination of the response scale categories. Measures of the internal consistency of each set of items measuring a construct and estimates of the precision of the instruments in consistently identifying person strata were also provided by the Rasch procedures and found to be acceptable.

The inability of confirmatory factor analysis (CFA) procedures to produce similar findings was an interesting, but not entirely unexpected result. The lack of congruence, however, did not call into question the construct validity, but rather provided further evidence for the need to carefully consider the analytic techniques used when evaluating instruments. Further research examining the appropriate use of both Rasch and CFA procedures is warranted.

The final study offered an assessment of criterion-related validity. As hypothesized, both efficacy and outcome measures were correlated with measures of effort. The pattern of correlations also offered evidence for the need to distinguish between efficacy and outcome expectations and to measure them independently. Other criterion measures, such as teacher planning and teacher engagement, were not correlated with measures of efficacy and outcome expectations as hypothesized. More well-defined variables and the use of multiple indicators of planning and engagement should be developed to further investigate these hypothesized relationships.

Future Research

One of the goals at the beginning of this project was to develop measures of teacher efficacy and teacher outcome expectations in line with Bandura' theory and recommendations for measurement. Toward this goal I intended to measure these constructs along his three dimensions of magnitude, strength, and generality. I obtained evidence of the first dimension, magnitude, through results indicating that the content of the items are measuring easy, moderately difficult, and difficult to endorse instructional activities. I obtained evidence for the second dimension, strength, through the distribution of teachers along the efficacy and outcome expectancy continuums and their appropriate use, in most cases, of the response scale categories. However, I was unable to assess the generality of teacher efficacy and outcome expectations due to the length of an instrument assessing both content domains and the resulting split of the initial instrument into two instruments for final administration.

Future studies investigating this dimension of generality in the context of teacher efficacy and outcome expectations should be conducted. It is important to determine whether elementary teachers' efficacy and outcome expectations are consistent across domains (reading, mathematics, and others) especially since many are required to teach multiple subjects at their assigned grade level. Other studies examining generality might investigate the generality of teachers' efficacy and outcome expectations in a single content domain but across the grade levels for which the teachers are certified. In either case, my expectations are that teachers will to some extent differentiate between their efficacy and outcome expectations for teaching different domains and different grade levels.

To facilitate the investigation of the generality dimension and to increase the ease of administering either instrument, I need to provide shorter versions of each instrument. The results of the Rasch analyses provide me with a process for doing this. It is possible to drop items from the instrument that are measuring the same level of difficulty as other items since these additional items are not adding much to the ability to discriminate among respondents' level of efficacy or outcome expectations. For example, in Table 9 and Figure 2 the items asking about teachers reading efficacy for integrating word processing activities, planning activities around student interests, and adjusting activities to students' skills have similar item measures and fall at the same point along the continuum of difficulty. Potentially, I could drop two of these three items from the instrument without altering the construct validity. Such a process could yield shorter instruments and increase the likelihood that teachers could respond to both instruments without experiencing fatigue. However, such a process would lead to a loss in the content validity of the instruments due to the instructional activities that would be absent from the assessment.

Other item modification studies are also the foci of future research. While some items might be dropped as suggested above to decrease instrument length, the Rasch results also show the need for construction and addition of more items to each scale. In all cases, the items defining each scale failed to assess the upper ends of the efficacy and outcome expectation continuums. This indicates that more items need to be developed that assess the expectations for instructional strategies that fall further up on the difficulty continuum if we want to be able to distinguish among many of the teachers reporting very high efficacy or outcome expectations. Interviews with practicing teachers might provide a method for obtaining further suggestions for instructional strategies that they find very difficult to implement and that should be candidates for item construction and evaluation.

Further investigation of the present items remaining in each scale for indication of item bias are also warranted. It is possible that teachers respond differentially to items depending on a variety of demographic measures. Potential biases could be whether teachers are from rural, urban, or suburban school districts, the number of students in their classrooms, or teacher gender. Examination of these potential biases would entail grouping the respondents along these categories, calibrating the items separately within these groups, and plotting item difficulty measures for each possible pair. If item bias exists, the plot would show excessive dispersion (i. e., ± 2 S.E.) away from the identity line. Responses for items identified as creating bias for or against certain groups of teachers could be treated as missing for this group or these items could be eliminated from the instrument.

Another focus of research involving possible instrument modification is the assessment of outcome expectations related to another outcome of interest. All instructional strategies included in the present instruments were assessed according to teachers' confidence that they contributed to the outcome of increased student achievement. While this is one of the outcome goals teachers possess for their students it is certainly not the only outcome teachers focus on in the classroom setting. Teachers often evaluate instructional strategies on their ability to motivate students to learn specific content or their ability to minimize classroom management problems. Modification of measures of outcome expectations in these content domains with a focus on a different outcome might provide additional information about teachers' beliefs that could enhance our ability to predict their behaviors.

A study should also be conducted to identify the optimal categorization for the rating scale used for these scales. In the present studies two of the four scales had a lack of order in their step calibration measures indicating that respondents were not using the response scale as intended. It may be that categories need to be collapsed or omitted from the response scale. Such an investigation would involve reanalyzing the present data through recoding into all possible category collapsings, and analyzing the Rasch results for each new rating scale through examination of person and item fit indices, average measures and step calibration measures, and person and item separation indices. Conclusions based on these analyses would indicate which rating response scale provides the best fit and reliability.

Finally, further evidence for criterion validity related to teachers' effort, persistence, and other teaching behaviors should be a focus of future validation studies. More well-defined measures of effort and other teaching behaviors such as planning and engagement in instruction could be developed for future investigations. Qualitative analyses of interviews with teachers and observations of teachers during classroom teaching might identify important or unique teaching behaviors or thinking indicative of teachers with low or high efficacy or outcome expectations.

Future Utility of the Instruments

The use of these instruments for teacher program planning and evaluation is a possible area of future research. Identifying those instructional strategies for which teachers possess low confidence in their ability to employ can be useful when designing staff development workshops. Such workshops should also provide teachers with evidence that these instructional strategies are likely to positively influence student achievement or another outcome of interest and importance such as student motivation. Such an emphasis is likely to increase teachers' outcome expectations for these instructional strategies while at the same time increasing their efficacy expectations.

Pre- and post- assessment of efficacy and outcome expectations can be used to identify areas for staff development and to assess the ability of the workshop to effect teacher change in efficacy or outcome expectations, and potentially their behaviors related to these instructional strategies. Teacher educators might also use this approach to planning programs and courses that will specifically address areas in which preservice teachers report low efficacy and/or outcome expectations.

As I mentioned at the beginning of this project, I developed the teacher efficacy expectations and teacher outcome expectations instruments within a specific context to meet the needs and purposes of continued research, specifically with a state database of student achievement in math and reading. Current theory and research on teacher efficacy suggests that it is quite possible that variations in some combination of teachers' efficacy and outcome expectations, and teachers' effort, persistence, and choice of instructional tasks could account for the reported variance in teachers' student gain scores in reading and math. These instruments can be used to examine the influences of efficacy and outcome expectations on teacher effort and subsequent student achievement in mathematics and reading. Future research utilizing this instrument also could include investigation and identification of variables that influence teacher efficacy expectations and teacher outcome expectations in the content domains of reading and math, and further examination of the influence of teacher efficacy expectations and teacher outcome expectations on teacher effort, persistence, and choice or preference for instructional tasks; and student motivation and achievement.

BIBLIOGRAPHY

Adams, R. J., & Wright, B. D. (1994). When does misfit make a difference? In M. Wilson (Ed.), Objective measurement: Theory into practice, Vol. 2. (pp. 244-270). Norwood, NJ: Ablex Publishing Corporation.

Anderson, R. N., Greene, M. L., & Loewen, P. S. (1988). Relationships among teachers' and students' thinking skills, sense of efficacy, and student achievement. The Alberta Journal of Educational Research, 34, 148-165.

Armor, D., Conry-Osequers, P., Cox, M., King, N., McDonnell, L., Pascal, A., Pauly, E., & Zellman, G. (1976). Analysis of the school preferred reading program in selected Los Angeles minority schools. (Report No. R-2007-LAUSD). Santa Monica, CA: The Rand Corporation. (ERIC Document Reproduction Service No. ED 130 243)

Ashton, P. T. (1984). Teacher efficacy: A motivational paradigm for effective teacher education. Journal of Teacher Education, 35 (5), 28-32.

Ashton, P. (1985). Motivation and the teacher's sense of efficacy. In C. Ames & R. Ames (Eds.), Research on motivation in education: The classroom milieu (pp. 141-174). Orlando, FL: Academic Press.

Ashton, P. T., Webb, R. B., & Doda, N. (1983). A study of teachers' sense of efficacy (Final Report). (Contract No. 400-79-0075)) Gainesville, FL: University of Florida. National Institute of Education. (ERIC Document Reproduction Service No. ED 231 833)

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

Bandura, A. (1986). Self-efficacy. In Social foundations of thought and action: A social cognitive theory, (pp. 390-453). Englewood Cliffs, NJ: Prentice-Hall.

Banerji, M., Smith, R. M., & Dedrick, R. F. (1997). Dimensionality of an early childhood scale using Rasch analysis and confirmatory factor analysis. Journal of Outcome Measurement, 1(1), 56-85.

Bentler, P. M. (1990). Comparative fit indexes in structural models. Psychological Bulletin, 107, 238-246.

Bentler, P. M. (1995). EQS: Structural Equations Program Manual. Encino, CA: Multivariate Software, Inc.

Curda, S. K., Curda, L. K., & Kleine, P. F. (1996, October). Developing a measure of preservice teachers' global and content specific efficacy. Paper presented at the annual convention of the Midwestern Educational Research Association, Chicago, IL.

De Ayala, R. J. (1993). An introduction to polytomous item response theory models. Measurement and Evaluation in Counseling and Development, 25, 172 - 189.

Denham, C. H., & Michael, J. J. (1981, Spring). Teacher sense of efficacy: A definition of the construct and a model for further research. Educational Research Quarterly, 39-63.

Forsyth, R., Saisangjan, U., & Gilmer, J. (1981). Some empirical results related to the robustness of the Rasch model. Applied Psychological Measurement, 5, 175-186.

Gable, R. K., & Wolf, M. B. (1993). Instrument development in the affective domain: Measuring attitudes and values in corporate and school settings (2nd ed.). Boston, MA: Kluwer Academic Publishers.

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

Guskey, T. R. (1981). Measurement of the responsibility teachers assume for academic success and failures in the classroom. Journal of Teacher Education, 32(3), 44-51.

Guskey, T. R. (1987). Context variables that affect measures of teacher efficacy. Journal of Educational Research, 81, 41-47.

Guskey, T. R. & Passaro, P. D. (1994). Teacher efficacy: A study of construct dimensions. American Educational Research Journal, 31, 627-643.

Hambleton, R. K., & Swaminathan, H. (1985). Item response theory: Principles and applications. Boston, MA: Kluwer Nijhoff Publishing.

Hambleton, R. K., & van der Linden, W. J. (1982). Advances in item response theory and applications: An introduction. Applied Psychological Measurement, 6, 373-378.

Jöreskog, K. G., & Sörbom, D. (1988). LISREL 7: A guide to the program and applications. Chicago: SPSS Inc.

Kilgore, K. M., Fisher, Jr., W. P., Silverstein, B., Harley, J. P., & Harvey, R. F. (1993). Application of Rasch analysis to the patient evaluation and conference system. In C. Granger & G. Gresham (Eds.), Physical Medicine and Rehabilitation Clinics of North America: New developments in functional assessment (pp. 493-515). Philadelphia: W. B. Sanders.

Lanier, J., & Little, J. (1986). Research on teacher education. In M. Wittrock (Ed.), Handbook of research on teaching (3rd ed.). New York: Macmillan.

Linacre, M., & Wright, B. D. (1995). BIGSTEPS computer program. Chicago: MESA Press

Linacre, M., & Wright, B. D. (1994). Chi-square fit statistics. Rasch Measurement Transactions, 8, 360-361.

Ludlow, L. H., Haley, S. M., & Gans, B. M. (1992). A hierarchical model of functional performance in rehabilitation medicine. Evaluation and Health Professions, 15(4), 59-74.

Lusardi, M. M., & Smith, Jr., E. V. (1997). Development of a scale to assess concern about falling and applications to treatment programs. Journal of Outcome Measurement, 1(1), 34-55.

Maddux, J. E., Norton, L. W., Stoltenberg, C. D. (1986). Self-efficacy expectancy, outcome expectancy, and outcome value: Relative effects on behavioral intentions. Journal of Personality and Social Psychology, 51, 783-789.

Manning, M. M., & Wright, T. L. (1983). Self-efficacy expectancies, outcome expectancies, and the persistence of pain control in childbirth. Journal of Personality and Social Psychology, 45, 421-431.

McArthur, D. L., Cohen, M. J., & Schandler, S. L. (1991). Rasch analysis of functional assessment scales: An example using pain behaviors. Archive of Physical and Medical Rehabilitation, 72, 296-304.

McKinley, R. L. An introduction to item response theory. Measurement and Evaluation in Counseling and Development, 22, 37-57.

Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Change in teacher efficacy and student self- and task-related beliefs in mathematics during the transition to junior high school. Journal of Educational Psychology, 81, 247-258.

Moore, W. P., & Esselman, M. E. (1992, April). Teacher efficacy, empowerment, and a focused instructional climate: Does student achievement benefit? Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. ED 350 252)

Mueller, R. O. (1996). Basic principles of structural equation modeling: An introduction to LISREL and EQS. New York: Springer.

O'Brien, M. L. (1992). Using Rasch procedures to understand psychometric structure in measures of personality. In M. Wilson (Ed.), Objective measurement: Theory into practice (pp. 61-76). Norwood, NJ: Ablex Publishing Corporation.

Owen, S. V. (1989, October). Building self-efficacy instruments. Paper presented at the annual Nursing Research Conference on Instrumentation in Nursing, Tucson, AZ.

Pajares, F. (1996b, April). Current directions in self research: Self-efficacy. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Pajares, F. (1996a, April). Self-efficacy as a motivational construct in academic settings. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Riggs, I. M., & Enochs, L. (1990). Toward the development of an efficacy belief instrument for elementary teachers. Science Education, 74, 625-637.

Roeser, R. W., Arbreton, A. J., & Anderman, E. M. (April, 1993). The relations among teacher beliefs and practices and student motivation across the school year. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Smith, R. M. (1996). A comparison of methods for determining dimensionality in Rasch measurement. Structural Equation Modeling, 3(1), 25-40.

Smith, R. M. (1986). Person fit in the Rasch model. Educational and Psychological Measurement, 46, 359-373.

Smith, R. M., & Miao, C. Y. (1994) Assessing unidimensionality for Rasch measurement. In M. Wilson (Ed.), Objective measurement: Theory into practice, Vol. 2. (pp. 316-327). Norwood, NJ: Ablex Publishing Corporation.

Smylie, M. (1990). Teacher efficacy at work. In P. Reyes (Ed.), Teachers and their workplace: Commitment, performance, and productivity. (pp. 48-66). Newbury Park, CA: Sage.

Stone, M. H., & Wright, B. D. (1988). Separation statistics in Rasch measurement. Research Memorandum No. 51. Chicago: MESA Psychometric Laboratory, The University of Chicago.

Tabachnick, B. G., & Fidell, L. S. (1996). Using multivariate statistics (3rd ed.). New York: Harper Collins College Publishers.

Tracz, S. M., & Gibson, S. (1986, November). Effects of efficacy on academic achievement. Paper presented at the annual meeting of the California Educational Research Association, Marina del Rey, CA. (ERIC Document Reproduction Service No. ED 281 853)

Vispoel, W. P., & Chen, P. (1990, April). Measuring self-efficacy: The state of the art. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Weiss, D. J., & Davison, M. L. (1981). Test theory and methods. Annual Review of Psychology , 32, 629-658.

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

Wright, B. D. (1994). A unidimensionality coefficient. Rasch Measurement Transactions, 8, 385.

Wright, B. D. (1996). Sample-free test calibration and person measurement. In A. W. Ward, H. W. Stoeker, & M. M. Ward (Eds.), Educational measurement: Origins, theories, and explications: Vol. 1. Basic concepts and theories (pp. 107-123). New York, NY: University Press of America, Inc.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. Journal of Educational Measurement, 14, 97-116.

Wright, B. D., & Linacre, M. (1994). Reasonable mean-square fit values. Rasch Measurement Transactions, 8, 370.

Wright, B. D., & Masters, G. N. (1982). Rating Scale Analysis. Chicago, IL: MESA Press.

Wright, B. D., & Stone, M. H. (1988a). Reliability in Rasch measurement. Research Memorandum No. 53. Chicago: MESA Psychometric Laboratory, The University of Chicago.

Wright, B. D., & Stone, M. H. (1988b). Validity in Rasch measurement. Research Memorandum No. 54. Chicago: MESA Psychometric Laboratory, The University of Chicago.

Zhu, W., Updyke, W., & Lewandowski, C. (1997, March). Optimal categorization of a rating scale: A longitudinal study. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Zimmerman, B. J. (1996, April). Misconceptions, problems, and dimensions in measuring self-efficacy. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

APPENDICES

Appendix A

Instrument Specifications for Teacher Efficacy and Outcome Expectancy Instrument

Instrument Specifications for Teacher Efficacy and Outcome Expectancy Instrument

Conceptual Definitions, Number of Items, and Sample Item for Each Category

1. *Reading Efficacy Expectations* are a teacher's confidence that he or she possesses the skills necessary to prepare materials, plan lessons, provide engaging instruction, motivate students, manage a classroom, integrate technology, and meet the learning needs of individual students in Reading. Between nine and twelve items should be used to assess this category with items intending to represent easy, moderately difficult, and difficult instructional tasks. A sample item is "I can provide individualized instruction in Reading for students based on their interests."

2. *Math Efficacy Expectations* are a teacher's confidence that he or she possesses the skills necessary to prepare materials, plan lessons, provide engaging instruction, motivate students, manage a classroom, integrate technology, and meet the learning needs of individual students in Math. Between nine and twelve items should be used to assess this category with items intending to represent easy, moderately difficult, and difficult instructional tasks. A sample item is "I have the ability to provide students with challenging seatwork in Math."

3. *Reading Outcome Expectations* are a teacher's confidence that the instructional activities and strategies in which he/she engages in the classroom will improve student achievement in Reading. Between nine and twelve items should be used to assess this category with items intending to represent easy, moderately difficult, and difficult instructional tasks. Each item should correspond with a reading efficacy expectation item. A sample item is "Providing individualized instruction in

Reading for students based on their interests will improve student achievement in Reading."

4. *Math Outcome Expectations* are a teacher's confidence that the instructional activities in which he/she engages in the classroom will improve student achievement in Math. Between nine and twelve items should be used to assess this category with items intending to represent easy, moderately difficult, and difficult instructional tasks. Each item should correspond with a math efficacy expectation item. A sample item is "Providing students with challenging seatwork will improve student achievement in Math."

Response Format

The instrument asks respondents to report their degree of certainty (a) in successfully performing the tasks described by the efficacy expectation items and (b) that the instructional activities described account for improved student achievement. Each item will be responded to using a Likert-type format with five anchors ranging from "1" to "5" and labeled with ranges of percent certainty, from "0 - 19% certain" to "80 - 100% certain."

Appendix B

Teacher Efficacy and Outcome Expectations Instrument

Teacher Efficacy and Outcome Expectations Instrument

Part I Directions: Teachers are often more or less confident they can successfully employ specific instructional methods or strategies with their students towards the goal of improving their achievement in Reading or Math. Using the scale provided, indicate your degree of confidence for successfully using the instructional methods or strategies described with the students you are presently teaching. For each item, circle the number that best represents your degree of confidence.

1	**2**	**3**	**4**	**5**
0 - 19% Confident	**20 - 39% Confident**	**40 - 59% Confident**	**60 - 79% Confident**	**80 - 100% Confident**

How confident am I that ...

1.	I have the skills to provide instruction in Reading based on students' interests.	1	2	3	4	5
2.	I am capable of integrating writing activities into my Reading instruction.	1	2	3	4	5
3.	I have the skills needed to adjust my classroom activities in Reading to the learning needs of individual students.	1	2	3	4	5
4.	I am capable of providing appropriate learning experiences in Reading for diverse learners (e.g., learning disabled, attention deficit, non-English speaking, gifted).	1	2	3	4	5
5.	I have the ability to adjust my classroom activities in Reading to account for differences in individual students' skills.	1	2	3	4	5
6.	I am capable of providing instruction in phonics skills when teaching Reading.	1	2	3	4	5
7.	I have the ability to integrate Reading activities into other curriculum areas.	1	2	3	4	5
8.	I have the ability to provide students with challenging seatwork in Reading.	1	2	3	4	5
9.	I have the ability to prepare engaging materials to teach Reading.	1	2	3	4	5
10.	I have the necessary skills for planning effective lessons in Reading.	1	2	3	4	5
11.	I have the ability to apply cooperative group learning strategies in teaching Reading.	1	2	3	4	5
12.	I am capable of providing students with the opportunity to choose their own literature books for Reading instruction.	1	2	3	4	5

1	2	3	4	5
0 - 19% Confident	**20 - 39% Confident**	**40 - 59% Confident**	**60 - 79% Confident**	**80 - 100% Confident**

How confident am I that ...

13.	I am capable of conducting whole group instruction in Reading.	1	2	3	4	5
14.	I am capable of conducting small group instruction in Reading.	1	2	3	4	5
15.	I have the skills to organize learning activities in Reading according to students' backgrounds.	1	2	3	4	5
16.	I have the skills to select appropriate computer software to integrate into Reading instruction.	1	2	3	4	5
17.	I am capable of providing learning activities for students that allow them to use computers or other technology during Reading instruction.	1	2	3	4	5
18.	I am capable of using computer tutorials to provide individualized instruction for students in Reading.	1	2	3	4	5
19.	I am capable of using CD-ROMs that contain children's literature books that match the goals of my Reading curriculum.	1	2	3	4	5
20.	I have the ability to incorporate word processing activities (e.g., journal or story writing) into my Reading instruction.	1	2	3	4	5
21.	I have the ability to plan activities for students to author multimedia projects in Reading (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage).	1	2	3	4	5
22.	I am capable of using drill and practice software to reach instructional goals in Reading.	1	2	3	4	5
23.	I have the skills to integrate the use of e-mail into relevant Reading activities.	1	2	3	4	5
24.	I have the ability to plan Reading lessons that allow students to use the World Wide Web to access relevant sources for curriculum related projects.	1	2	3	4	5
25.	I have the skills to provide instruction in Math based on my students' interests.	1	2	3	4	5
26.	I am capable of providing Math instruction using a variety of hands-on manipulatives.	1	2	3	4	5
27.	I have the skills needed to adjust my classroom activities in Math to the learning needs of individual students.	1	2	3	4	5

1	2	3	4	5
0 - 19% Confident	**20 - 39% Confident**	**40 - 59% Confident**	**60 - 79% Confident**	**80 - 100% Confident**

How confident am I that ...

28.	I am capable of providing appropriate learning experiences in Math for diverse learners (e.g., learning disabled, attention deficit, non-English speaking, gifted).	1	2	3	4	5
29.	I have the ability to adjust my classroom activities in Math to account for differences in individual students' skills.	1	2	3	4	5
30.	I am capable of providing timed practice tests for students during Math instruction.	1	2	3	4	5
31.	I have the ability to integrate Math activities into other curriculum areas.	1	2	3	4	5
32.	I have the ability to provide students with challenging seatwork in Math.	1	2	3	4	5
33.	I have the ability to prepare engaging materials to teach Math.	1	2	3	4	5
34.	I have the necessary skills for planning effective lessons in Math.	1	2	3	4	5
35.	I have the ability to apply cooperative group learning strategies in teaching Math.	1	2	3	4	5
36.	I have the ability to provide worked examples during Math instruction.	1	2	3	4	5
37.	I am capable of conducting whole group instruction in Math.	1	2	3	4	5
38.	I am capable of conducting small group instruction in Math.	1	2	3	4	5
39.	I have the skills to organize learning activities in Math according to students' backgrounds.	1	2	3	4	5
40.	I have the skills to select appropriate computer software to integrate into my Math curriculum.	1	2	3	4	5
41.	I am capable of providing learning activities for students that allow them to use computers or other technology to attain an instructional goal in Math.	1	2	3	4	5
42.	I am capable of using computer tutorials to provide individualized Math instruction for students.	1	2	3	4	5
43.	I have the ability to use computer games that match the goals of my Math curriculum.	1	2	3	4	5

1	2	3	4	5
0 - 19% Confident	20 - 39% Confident	40 - 59% Confident	60 - 79% Confident	80 - 100% Confident

How confident am I that ...

44.	I have the ability to plan activities for students to author multimedia projects in Math (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage).	1	2	3	4	5
45.	I am capable of using drill and practice software to reach instructional goals in Math.	1	2	3	4	5
46.	I have the ability to incorporate spreadsheet activities into my Math instruction.	1	2	3	4	5
47.	I am capable of planning lessons that use videodiscs to reach an instructional goal in Math.	1	2	3	4	5
48.	I have the skills to plan lessons that allow students to use the World Wide Web to access relevant sources for curriculum related projects in Math.	1	2	3	4	5

Part II Directions: Teachers are often more or less confident that successful implementation of specific instructional methods or strategies are effective in reaching instructional goals they have for their students. Using the scale provided, indicate your degree of confidence that the specific instructional method or strategy described will be **effective** in improving either the standardized Reading or Math achievement of the students you are presently teaching. For each item, circle the number that best represents your degree of confidence.

1	2	3	4	5
0 - 19% Confident	20 - 39% Confident	40 - 59% Confident	60 - 79% Confident	80 - 100% Confident

When implemented appropriately, how confident are you that an effective strategy for improving students' standardized reading achievement scores is ...

1.	providing instruction based on students' interests	1	2	3	4	5
2.	integrating writing activities into the Reading instruction.	1	2	3	4	5
3.	adjusting classroom activities to the learning needs of individual students.	1	2	3	4	5
4.	providing appropriate learning experiences for diverse learners (e.g., learning disabled, attention deficit, non-English speaking, gifted).	1	2	3	4	5
5.	adjusting classroom activities to account for differences in individual students' Reading skills.	1	2	3	4	5

1	2	3	4	5
0 - 19% Confident	20 - 39% Confident	40 - 59% Confident	60 - 79% Confident	80 - 100% Confident

When implemented appropriately, how confident are you that an effective strategy for improving students' standardized reading achievement scores is ...

6.	providing instruction in phonics skills.	1	2	3	4	5
7.	integrating Reading activities into other curriculum areas.	1	2	3	4	5
8.	providing students with challenging seatwork.	1	2	3	4	5
9.	preparing engaging materials to teach Reading.	1	2	3	4	5
10.	planning effective lessons in Reading.	1	2	3	4	5
11.	applying cooperative group learning strategies.	1	2	3	4	5
12.	providing students with the opportunity to choose their own literature books for reading.	1	2	3	4	5
13.	conducting whole group instruction.	1	2	3	4	5
14.	conducting small group instruction.	1	2	3	4	5
15.	organizing learning activities according to students' backgrounds.	1	2	3	4	5
16.	selecting appropriate computer software to integrate into the instruction.	1	2	3	4	5
17.	providing learning activities for students that allow them to use computers or other technology during instruction.	1	2	3	4	5
18.	using computer tutorials to provide individualized instruction for students in Reading.	1	2	3	4	5
19.	using CD-ROMs of children's literature that match the goals of the Reading curriculum.	1	2	3	4	5
20.	incorporating word processing activities (e.g., journal or story writing) into the instruction.	1	2	3	4	5
21.	planning activities for students to author multimedia projects in Reading (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage).	1	2	3	4	5
22.	using drill and practice software to reach instructional goals in Reading.	1	2	3	4	5
23.	integrating the use of e-mail into relevant Reading activities.	1	2	3	4	5

1	2	3	4	5
0 - 19% Confident	**20 - 39% Confident**	**40 - 59% Confident**	**60 - 79% Confident**	**80 - 100% Confident**

When implemented appropriately, how confident are you that an effective strategy for improving students' standardized reading achievement scores is ...

24.	planning lessons that allow students to use the World Wide Web to access relevant sources for curriculum related projects.	1	2	3	4	5

When implemented appropriately, how confident are you that an effective strategy for improving students' standardized math achievement scores is ...

25.	providing instruction based on my students' interests.	1	2	3	4	5
26.	providing instruction using a variety of hands-on manipulatives.	1	2	3	4	5
27.	adjusting classroom activities in Math to the learning needs of individual students.	1	2	3	4	5
28.	providing appropriate learning experiences for diverse learners (e.g., learning disabled, attention deficit, non-English speaking, gifted).	1	2	3	4	5
29.	adjusting classroom activities to account for differences in individual students' Math skills.	1	2	3	4	5
30.	providing timed practice tests for students during instruction.	1	2	3	4	5
31.	integrating Math activities into other curriculum areas.	1	2	3	4	5
32.	providing students with challenging seatwork in Math.	1	2	3	4	5
33.	preparing engaging materials to teach Math.	1	2	3	4	5
34.	planning effective lessons in Math.	1	2	3	4	5
35.	applying cooperative group learning strategies.	1	2	3	4	5
36.	providing worked examples during instruction.	1	2	3	4	5
37.	conducting whole group instruction.	1	2	3	4	5
38.	conducting small group instruction.	1	2	3	4	5
39.	organizing learning activities according to students' backgrounds.	1	2	3	4	5
40.	selecting appropriate computer software to integrate into the Math curriculum.	1	2	3	4	5

1	2	3	4	5
0 - 19% Confident	**20 - 39% Confident**	**40 - 59% Confident**	**60 - 79% Confident**	**80 - 100% Confident**

When implemented appropriately, how confident are you that an effective strategy for improving students' standardized math achievement scores is ...

41.	providing learning activities for students that allow them to use computers or other technology to attain an instructional goal in Math.	1	2	3	4	5
42.	using computer tutorials to provide individualized Math instruction for students.	1	2	3	4	5
43.	using computer games that match the goals of the Math curriculum.	1	2	3	4	5
44.	planning activities for students to author multimedia projects in Math (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage).	1	2	3	4	5
45.	using drill and practice software to reach instructional goals in Math.	1	2	3	4	5
46.	incorporating spreadsheet activities into Math instruction.	1	2	3	4	5
47.	planning lessons that use videodiscs to reach an instructional goal in Math.	1	2	3	4	5
48.	planning lessons that allow students to use the World Wide Web to access relevant sources for curriculum related projects in Math.	1	2	3	4	5

Appendix C

Content Expert Rating Form for Conceptual and Operational Definitions

Content Expert Rating Form for Conceptual and Operational Definitions

Directions: Please review the summary of the background theory, the teacher efficacy and outcome expectation instrument, and the proposed use of the instrument provided below. Then rate and comment on the areas listed below. Use the rating scale below for each of the areas.

Poor					Excellent
1	2	3	4	5	6

1. Conceptual definition of reading efficacy expectations RATING _____
2. Conceptual definition of math efficacy expectations RATING _____

Comments:

3. Correspondence between the conceptual definition of reading efficacy expectations and the instrument items (PART I: 1-24). RATING _____
4. Correspondence between the conceptual definition of math efficacy expectations and the instrument items (PART I: 25-48). RATING _____

Comments:

5. Conceptual definition of reading outcome expectations RATING _____
6. Conceptual definition of math outcome expectations RATING _____

Comments:

7. Correspondence between the conceptual definition of reading outcome expectations and the instrument items (PART II: 1-24). RATING _____

8. Correspondence between the conceptual definition of math outcome expectations and the instrument items (PART II: 25-48). RATING _____

Comments:

9. Adequacy of the definitions as they relate to the proposed use of the instrument RATING _____

10. Adequacy of the instrument items as they relate to the proposed use of the instrument RATING _____

Comments:

Brief Theoretical Background and Purpose of Instrument

Bandura

Bandura described outcome expectations and efficacy expectations as instrumental in accounting for behavioral and motivational changes in behavior. Outcome expectations are beliefs that behaviors will or will not lead to desirable outcomes. Efficacy expectations are beliefs that one does or does not possess the required skills to bring about the performance and are often referred to as self-efficacy.

The Current Study

The current study proposes to develop and validate a measure of teacher efficacy expectations and teacher outcome expectations specific to the needs and purposes of continued research investigating and identifying variables that influence teacher efficacy and outcome expectations in the domains of math and reading and also discovering if such measures can reliably predict or account for significant variance in teacher effort, persistence, and choice or preference for instructional tasks. This study intends to remedy early empirical research flaws in the measurement of teacher efficacy and outcome expectations through the construction and validation of a measure more clearly in alignment with Bandura's theory.

Context for the Study

This current study will be constructing and validating the teacher efficacy and outcome expectations instrument within a specific context. As with many states, there is mandatory testing of students from grades 2 - 12 in the state of Tennessee. Presently, this researcher has access to the database which provides an assessment of teachers' past and present achievements through measures of their students' gain scores in reading and math. While analysis of this data is ongoing and findings are interesting, no external data has been collected to account for many of the findings. For example, some teachers consistently have high student gains each year in reading or math while others do not (even when controlling for school setting, size, ethnic composition, and a variety of other variables). Current theory and research on teacher efficacy suggests that it is quite possible that variations in some combination of teachers' efficacy expectations, outcome expectations, and teachers' effort, persistence, and choice of instructional tasks can account for the reported variance in teachers' student gain scores in reading and math. However, before we can confidently draw conclusions from an investigation testing these ideas we need to have confidence that the measures of teacher efficacy expectations and outcome expectations for the domains of interest are valid and reliable.

Conceptual Definitions for Teacher Efficacy and Outcome Expectancy Instrument

1. *Reading Efficacy Expectations* are a teacher's confidence that he or she possesses the skills necessary to successfully implement a variety of instructional methods or strategies for teaching Reading. Items should represent easy, moderately difficult, and difficult instructional tasks to implement.

2. *Math Efficacy Expectations* are a teacher's confidence that he or she possesses the skills necessary to successfully implement a variety of instructional methods or strategies for teaching Math. Items should represent easy, moderately difficult, and difficult instructional tasks to implement.

3. *Reading Outcome Expectations* are a teacher's confidence that specific instructional methods or strategies implemented successfully during Reading instruction are effective ways to teach Reading. Each item is constructed to align with an item measuring Reading Efficacy Expectations.

4. *Math Outcome Expectations* are a teacher's confidence that specific instructional methods or strategies implemented successfully during Math instruction are effective ways to teach Math. Each item is constructed to align with an item measuring Math Efficacy Expectations.

Response Format

The instrument asks respondents to report their degree of confidence (a) in successfully implementing the instructional methods or strategies used when teaching Reading or Math and (b) that the instructional methods or strategies are effective for teaching Reading or Math. Each item will be responded to using a Likert-type format with five anchors ranging from "1" to "5" and labeled with ranges of percent confidence, from "0 - 19 % confidence" to "80 - 100% confidence."

Validation

The responses to the items measuring one of the four possible categories will be used to conduct four separate Rasch rating scale analyses (one for each construct) and confirmatory factor analysis for construct validity purposes. The Rasch person ability measures for each group of items will also be used to conduct correlational analyses with measures of teacher effort, persistence, and preference for instructional task for criterion-related tests of validity. Additional questions assessing effort, persistence, and preference related to technology are included in the event that the technology-related items form separate constructs. The desirable final pool of items for each domain of efficacy is lower than the number of items presently used in the instrument. Instrument development procedures suggest to begin with more than the number of items you wish to have in your final pool. Most likely, some items will be dropped at each stage of the validity and reliability analyses. Once the final pool has been reached final reliability and validity estimates will be reported.

Appendix D

Content Validity Category Rating Form

Content Validity Category Rating Form

Directions: The statements that follow are being considered for inclusion in a survey of teacher efficacy and outcome expectations. Please assist me in reviewing the content of the statements by providing two ratings for each statement. The conceptual definitions of the categories these statements are supposed to reflect as well as the rating instructions are below.

Categories	Conceptual Definition
I. Reading Efficacy Expectations	An assessment of a teacher's confidence that he or she possesses the skills necessary to successfully implement a variety of instructional methods or strategies for teaching Reading.
II. Math Efficacy Expectations	An assessment of a teacher's confidence that he or she possesses the skills necessary to successfully implement a variety of instructional methods or strategies for teaching Math.
III. Reading Outcome Expectations	An assessment of a teacher's confidence that specific instructional methods or strategies implemented successfully during Reading instruction are effective strategies for improving students' standardized reading achievement scores.
IV. Math Outcome Expectations	An assessment of a teacher's confidence that specific instructional methods or strategies implemented successfully during Math instruction are effective strategies for improving students' standardized math achievement scores.

Rating Tasks

A. Please indicate the category that each statement best fits by circling the appropriate category numeral. (Statements not fitting into any category should be placed in Category V.)

B. Please indicate how strongly you feel about your placement of the statement into the category by circling the appropriate number as follows:

4 absolutely sure
3 very sure
2 sure
1 not very sure

How confident am I that ...	Categories					Rating			
1. I have the skills to provide instruction in Reading based on students' interests.	I	II	III	IV	V	1	2	3	4
2. I am capable of integrating writing activities into my Reading instruction.	I	II	III	IV	V	1	2	3	4
3. I have the skills needed to adjust my classroom activities in Reading to the learning needs of individual students.	I	II	III	IV	V	1	2	3	4
4. I am capable of providing appropriate learning experiences in Reading for diverse learners (e.g., learning disabled, attention deficit, non-English speaking, gifted).	I	II	III	IV	V	1	2	3	4
5. I have the ability to adjust my classroom activities in Reading to account for differences in individual students' skills.	I	II	III	IV	V	1	2	3	4
6. I am capable of providing instruction in phonics skills when teaching Reading.	I	II	III	IV	V	1	2	3	4
7. I have the ability to integrate Reading activities into other curriculum areas.	I	II	III	IV	V	1	2	3	4
8. I have the ability to provide students with challenging seatwork in Reading.	I	II	III	IV	V	1	2	3	4
9. I have the ability to prepare engaging materials to teach Reading.	I	II	III	IV	V	1	2	3	4
10. I have the necessary skills for planning effective lessons in Reading.	I	II	III	IV	V	1	2	3	4
11. I have the ability to apply cooperative group learning strategies in teaching Reading.	I	II	III	IV	V	1	2	3	4
12. I am capable of providing students with the opportunity to choose their own literature books for Reading instruction.	I	II	III	IV	V	1	2	3	4
13. I am capable of conducting whole group instruction in Reading.	I	II	III	IV	V	1	2	3	4
14. I am capable of conducting small group instruction in Reading.	I	II	III	IV	V	1	2	3	4
15. I have the skills to organize learning activities in Reading according to students' backgrounds.	I	II	III	IV	V	1	2	3	4

How confident am I that ...	Categories					Rating			
16. I have the skills to select appropriate computer software to integrate into Reading instruction.	I	II	III	IV	V	1	2	3	4
17. I am capable of providing learning activities for students that allow them to use computers or other technology during Reading instruction.	I	II	III	IV	V	1	2	3	4
18. I am capable of using computer tutorials to provide individualized instruction for students in Reading.	I	II	III	IV	V	1	2	3	4
19. I am capable of using CD-ROMs that contain children's literature books that match the goals of my Reading curriculum.	I	II	III	IV	V	1	2	3	4
20. I have the ability to incorporate word processing activities (e.g., journal or story writing) into my Reading instruction.	I	II	III	IV	V	1	2	3	4
21. I have the ability to plan activities for students to author multimedia projects in Reading (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage).	I	II	III	IV	V	1	2	3	4
22. I am capable of using drill and practice software to reach instructional goals in Reading.	I	II	III	IV	V	1	2	3	4
23. I have the skills to integrate the use of e-mail into relevant Reading activities.	I	II	III	IV	V	1	2	3	4
24. I have the ability to plan Reading lessons that allow students to use the World Wide Web to access relevant sources for curriculum related projects.	I	II	III	IV	V	1	2	3	4
25. I have the skills to provide instruction in Math based on my students' interests.	I	II	III	IV	V	1	2	3	4
26. I am capable of providing Math instruction using a variety of hands-on manipulatives.	I	II	III	IV	V	1	2	3	4
27. I have the skills needed to adjust my classroom activities in Math to the learning needs of individual students.	I	II	III	IV	V	1	2	3	4
28. I am capable of providing appropriate learning experiences in Math for diverse learners (e.g., learning disabled, attention deficit, non-English speaking, gifted).	I	II	III	IV	V	1	2	3	4

How confident am I that ...	Categories					Rating			
29. I have the ability to adjust my classroom activities in Math to account for differences in individual students' skills.	I	II	III	IV	V	1	2	3	4
30. I am capable of providing timed practice tests for students during Math instruction.	I	II	III	IV	V	1	2	3	4
31. I have the ability to integrate Math activities into other curriculum areas.	I	II	III	IV	V	1	2	3	4
32. I have the ability to provide students with challenging seatwork in Math.	I	II	III	IV	V	1	2	3	4
33. I have the ability to prepare engaging materials to teach Math.	I	II	III	IV	V	1	2	3	4
34. I have the necessary skills for planning effective lessons in Math.	I	II	III	IV	V	1	2	3	4
35. I have the ability to apply cooperative group learning strategies in teaching Math.	I	II	III	IV	V	1	2	3	4
36. I have the ability to provide worked examples during Math instruction.	I	II	III	IV	V	1	2	3	4
37. I am capable of conducting whole group instruction in Math.	I	II	III	IV	V	1	2	3	4
38. I am capable of conducting small group instruction in Math.	I	II	III	IV	V	1	2	3	4
39. I have the skills to organize learning activities in Math according to students' backgrounds.	I	II	III	IV	V	1	2	3	4
40. I have the skills to select appropriate computer software to integrate into my Math curriculum.	I	II	III	IV	V	1	2	3	4
41. I am capable of providing learning activities for students that allow them to use computers or other technology to attain an instructional goal in Math.	I	II	III	IV	V	1	2	3	4
42. I am capable of using computer tutorials to provide individualized Math instruction for students.	I	II	III	IV	V	1	2	3	4
43. I have the ability to use computer games that match the goals of my Math curriculum.	I	II	III	IV	V	1	2	3	4

How confident am I that ...	Categories	Rating
44. I have the ability to plan activities for students to author multimedia projects in Math (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage).	I II III IV V	1 2 3 4
45. I am capable of using drill and practice software to reach instructional goals in Math.	I II III IV V	1 2 3 4
46. I have the ability to incorporate spreadsheet activities into my Math instruction.	I II III IV V	1 2 3 4
47. I am capable of planning lessons that use videodiscs to reach an instructional goal in Math.	I II III IV V	1 2 3 4
48. I have the skills to plan lessons that allow students to use the World Wide Web to access relevant sources for curriculum related projects in Math.	I II III IV V	1 2 3 4

When implemented appropriately, how confident are you that an effective strategy for improving students' standardized reading achievement scores is ...	Categories	Rating
1. providing instruction based on students' interests	I II III IV V	1 2 3 4
2. integrating writing activities into the Reading instruction.	I II III IV V	1 2 3 4
3. adjusting classroom activities to the learning needs of individual students.	I II III IV V	1 2 3 4
4. providing appropriate learning experiences for diverse learners (e.g., learning disabled, attention deficit, non-English speaking, gifted).	I II III IV V	1 2 3 4
5. adjusting classroom activities to account for differences in individual students' Reading skills.	I II III IV V	1 2 3 4
6. providing instruction in phonics skills.	I II III IV V	1 2 3 4
7. integrating Reading activities into other curriculum areas.	I II III IV V	1 2 3 4
8. providing students with challenging seatwork.	I II III IV V	1 2 3 4
9. preparing engaging materials to teach Reading.	I II III IV V	1 2 3 4
10. planning effective lessons in Reading.	I II III IV V	1 2 3 4
11. applying cooperative group learning strategies.	I II III IV V	1 2 3 4

When implemented appropriately, how confident are you that an effective strategy for improving students' standardized reading achievement scores is ...

		Categories	Rating
12.	providing students with the opportunity to choose their own literature books for reading.	I II III IV V	1 2 3 4
13.	conducting whole group instruction.	I II III IV V	1 2 3 4
14.	conducting small group instruction.	I II III IV V	1 2 3 4
15.	organizing learning activities according to students' backgrounds.	I II III IV V	1 2 3 4
16.	selecting appropriate computer software to integrate into the instruction.	I II III IV V	1 2 3 4
17.	providing learning activities for students that allow them to use computers or other technology during instruction.	I II III IV V	1 2 3 4
18.	using computer tutorials to provide individualized instruction for students in Reading.	I II III IV V	1 2 3 4
19.	using CD-ROMs of children's literature that match the goals of the Reading curriculum.	I II III IV V	1 2 3 4
20.	incorporating word processing activities (e.g., journal or story writing) into the instruction.	I II III IV V	1 2 3 4
21.	planning activities for students to author multimedia projects in Reading (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage).	I II III IV V	1 2 3 4
22.	using drill and practice software to reach instructional goals in Reading.	I II III IV V	1 2 3 4
23.	integrating the use of e-mail into relevant Reading activities.	I II III IV V	1 2 3 4
24.	planning lessons that allow students to use the World Wide Web to access relevant sources for curriculum related projects.	I II III IV V	1 2 3 4

When implemented appropriately, how confident are you that an effective strategy for improving students' standardized math achievement scores is ...

		Categories	Rating
25.	providing instruction based on students' interests	I II III IV V	1 2 3 4

When implemented appropriately, how confident are you that an effective strategy for improving students' standardized math achievement scores is ...	Categories	Rating
26. providing instruction using a variety of hands-on manipulatives.	I II III IV V	1 2 3 4
27. adjusting classroom activities in Math to the learning needs of individual students.	I II III IV V	1 2 3 4
28. providing appropriate learning experiences for diverse learners (e.g., learning disabled, attention deficit, non-English speaking, gifted).	I II III IV V	1 2 3 4
29. adjusting classroom activities to account for differences in individual students' Math skills.	I II III IV V	1 2 3 4
30. providing timed practice tests for students during instruction.	I II III IV V	1 2 3 4
31. integrating Math activities into other curriculum areas.	I II III IV V	1 2 3 4
32. providing students with challenging seatwork in Math.	I II III IV V	1 2 3 4
33. preparing engaging materials to teach Math.	I II III IV V	1 2 3 4
34. planning effective lessons in Math.	I II III IV V	1 2 3 4
35. applying cooperative group learning strategies.	I II III IV V	1 2 3 4
36. providing worked examples during instruction.	I II III IV V	1 2 3 4
37. conducting whole group instruction.	I II III IV V	1 2 3 4
38. conducting small group instruction.	I II III IV V	1 2 3 4
39. organizing learning activities according to students' backgrounds.	I II III IV V	1 2 3 4
40. selecting appropriate computer software to integrate into the Math curriculum.	I II III IV V	1 2 3 4
41. providing learning activities for students that allow them to use computers or other technology to attain an instructional goal in Math.	I II III IV V	1 2 3 4
42. using computer tutorials to provide individualized Math instruction for students.	I II III IV V	1 2 3 4

When implemented appropriately, how confident are you that an effective strategy for improving students' standardized math achievement scores is ...	**Categories**					**Rating**			
43. using computer games that match the goals of the Math curriculum.	I	II	III	IV	V	1	2	3	4
44. planning activities for students to author multimedia projects in Math (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage).	I	II	III	IV	V	1	2	3	4
45. using drill and practice software to reach instructional goals in Math.	I	II	III	IV	V	1	2	3	4
46. incorporating spreadsheet activities into Math instruction.	I	II	III	IV	V	1	2	3	4
47. planning lessons that use videodiscs to reach an instructional goal in Math.	I	II	III	IV	V	1	2	3	4
48. planning lessons that allow students to use the World Wide Web to access relevant sources for curriculum related projects in Math.	I	II	III	IV	V	1	2	3	4

Appendix E

Content Validity Item Difficulty Rating Form

Content Validity Item Difficulty Rating Form

Part I Directions: The instructional activities below are being considered for inclusion in a teacher efficacy survey. Please assist me in reviewing these instructional activities by rating the difficulty of successfully completing each task. Use the scale provided to indicate your beliefs about the difficulty the average teacher might have implementing each instructional strategy effectively.

very easy	easy	avg	difficult	very difficult
1	**2**	**3**	**4**	**5**

1. Instruction in Reading based on students' interests	1	2	3	4	5
2. Integrating writing activities into Reading instruction	1	2	3	4	5
3. Adjusting activities in Reading to the learning needs of individual students	1	2	3	4	5
4. Providing appropriate Reading learning activities for diverse learners	1	2	3	4	5
5. Adjusting activities in Reading to account for differences in individual students' skills	1	2	3	4	5
6. Providing instruction in phonics skills when teaching Reading	1	2	3	4	5
7. Integrating Reading activities into other curriculum areas	1	2	3	4	5
8. Providing challenging seatwork in Reading	1	2	3	4	5
9. Preparing engaging materials in Reading	1	2	3	4	5
10. Planning effective lessons in Reading	1	2	3	4	5
11. Applying cooperative group learning strategies in Reading	1	2	3	4	5
12. Providing students with the opportunity to choose their own literature books for Reading instruction	1	2	3	4	5
13. Conducting whole group instruction in Reading	1	2	3	4	5
14. Conducting small group instruction in Reading	1	2	3	4	5

	very easy	easy	avg	difficult	very difficult
	1	2	3	4	5

15. Organizing activities in Reading according to students' background	1	2	3	4	5
16. Selecting appropriate computer software to integrate into Reading instruction	1	2	3	4	5
17. Providing activities that allow students to use technology to attain an instructional goal in Reading	1	2	3	4	5
18. Using computer tutorials to provide individualized instruction in Reading	1	2	3	4	5
19. Using CD-ROMs that contain children's literature books that match the goals of the Reading curriculum	1	2	3	4	5
20. Incorporating word processing activities into Reading instruction	1	2	3	4	5
21. Planning activities for students to author multimedia projects in Reading (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage)	1	2	3	4	5
22. Using drill and practice software to reach instructional goals in Reading	1	2	3	4	5
23. Integrating the use of e-mail into relevant Reading activities	1	2	3	4	5
24. Planning Reading lessons that allow students to use the World Wide Web to access relevant sources for curriculum related projects	1	2	3	4	5

All of the above instructional activities were sampled from the variety of instructional tasks or strategies teachers may use during Reading instruction. How well do you believe these instructional tasks represent the range of teachers' instructional practices in Reading.

1	2	3	4	5	6
Poor					Excellent

In addition to the activities listed above, can you think of two or three additional instructional activities that you or other teachers frequently use that you believe improves student achievement or aids you in reaching your instructional goals in Reading? ______________________________

Part II Directions: The instructional activities below are being considered for inclusion in a teacher efficacy survey. Please assist me in reviewing these instructional activities by rating the difficulty of successfully completing each task. Use the scale provided to indicate your beliefs about the difficulty the average teacher might have implementing each instructional strategy effectively.

very easy	easy	avg	difficult	very difficult
1	**2**	**3**	**4**	**5**

1. Instruction in Math based on students' interests	1	2	3	4	5
2. Providing Math instruction using a variety of hands-on manipulatives	1	2	3	4	5
3. Adjusting activities in Math to the learning needs of individual students	1	2	3	4	5
4. Providing appropriate Math learning activities for diverse learners	1	2	3	4	5
5. Adjusting activities in Math to account for differences in students' skills	1	2	3	4	5
6. Providing times practice tests for student during Math instruction	1	2	3	4	5
7. Integrating Math activities into other curriculum areas	1	2	3	4	5
8. Providing challenging seatwork in Math	1	2	3	4	5
9. Preparing engaging materials in Math	1	2	3	4	5
10. Planning effective lessons in Math	1	2	3	4	5
11. Applying cooperative group learning strategies in Math	1	2	3	4	5
12. Providing worked examples during Math instruction	1	2	3	4	5
13. Conducting whole group instruction in Math	1	2	3	4	5
14. Conducting small group instruction in Math	1	2	3	4	5
15. Organizing activities in Math according to students' background	1	2	3	4	5

	very easy	easy	avg	difficult	very difficult
	1	**2**	**3**	**4**	**5**
16. Selecting appropriate computer software to integrate into Math instruction	1	2	3	4	5
17. Providing activities that allow students to use technology to attain an instructional goal in Math	1	2	3	4	5
18. Using computer tutorials to provide individualized instruction in Math	1	2	3	4	5
19. Using computer games that match the goals of the Math curriculum	1	2	3	4	5
20. Planning lessons that use videodiscs for instructional goals in Math	1	2	3	4	5
21. Planning activities for students to author multimedia projects in Math (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage)	1	2	3	4	5
22. Using drill and practice software to reach instructional goals in Math	1	2	3	4	5
23. Incorporating spreadsheet activities into Math instruction	1	2	3	4	5
24. Planning Math lessons that allow students to use the World Wide Web to access relevant sources for curriculum related projects	1	2	3	4	5

All of the above instructional activities were sampled from the variety of instructional tasks or strategies teachers may use during Math instruction. How well do you believe these instructional tasks represent the range of teachers' instructional practices in Math.

1	2	3	4	5	6
Poor					Excellent

In addition to the activities listed above, can you think of two or three additional instructional activities that you or other teachers frequently use that you believe improves student achievement or aids you in reaching your instructional goals in Math? ____

Appendix F

Teacher Efficacy and Outcome Expectations Instrument (Pilot)

Teacher Efficacy and Outcome Expectations Instrument (Pilot)

Part I Directions: Teachers are often more or less confident they can successfully employ specific instructional methods or strategies with their students towards the goal of improving their achievement in Reading or Math. Using the scale provided, indicate your degree of confidence for successfully using the instructional methods or strategies described with the students you are presently teaching. As with many teachers, you may not actively employ some of the strategies listed below. For example, you may be using a whole language approach to teach reading or you may not have access to certain types of technology; however, you should still be able to report your confidence that you could successfully use a basal reader or apply the technology if you had access. Remember that the focus of these questions is your confidence in using the instructional strategies. These questions are not asking if you are presently using these strategies or agree with these strategies. For each item, circle the number that best represents your degree of confidence.

1	2	3	4	5
0 - 19% Confident	**20 - 39% Confident**	**40 - 59% Confident**	**60 - 79% Confident**	**80 - 100% Confident**

How confident am I that ...

1.	I am capable of providing instruction in Reading based on students' interests.	1	2	3	4	5
2.	I am capable of integrating writing activities into my Reading instruction.	1	2	3	4	5
3.	I have the ability to adjust my classroom activities in Reading to the learning needs of individual students.	1	2	3	4	5
4.	I am capable of providing appropriate learning experiences in Reading for diverse learners (e.g., learning disabled, attention deficit, non-English speaking, gifted).	1	2	3	4	5
5.	I have the ability to adjust my classroom activities in Reading to account for differences in individual students' skills.	1	2	3	4	5
6.	I am capable of providing instruction in phonics skills when teaching Reading.	1	2	3	4	5
7.	I have the ability to integrate Reading activities into other curriculum areas.	1	2	3	4	5
8.	I have the ability to provide students with challenging seatwork in Reading.	1	2	3	4	5
9.	I have the ability to prepare engaging materials to teach Reading.	1	2	3	4	5

1	2	3	4	5
0 - 19% Confident	20 - 39% Confident	40 - 59% Confident	60 - 79% Confident	80 - 100% Confident

How confident am I that ...

10.	I have the ability to plan effective lessons in Reading.	1	2	3	4	5
11.	I have the ability to apply cooperative group learning strategies in teaching Reading.	1	2	3	4	5
12.	I am capable of providing students with the opportunity to choose their own literature books for Reading instruction.	1	2	3	4	5
13.	I am capable of conducting whole group instruction in Reading.	1	2	3	4	5
14.	I am capable of conducting small group instruction in Reading.	1	2	3	4	5
15.	I am capable of organizing learning activities in Reading according to students' backgrounds.	1	2	3	4	5
16.	I am capable of selecting well-designed computer software to integrate into Reading instruction.	1	2	3	4	5
17.	I am capable of providing learning activities for students that allow them to use computers or other technology during Reading instruction.	1	2	3	4	5
18.	I am capable of using computer tutorials to provide individualized instruction for students in Reading.	1	2	3	4	5
19.	I am capable of using CD-ROMs that contain children's literature books that match the goals of my Reading curriculum.	1	2	3	4	5
20.	I have the ability to incorporate word processing activities (e.g., journal or story writing) into my Reading instruction.	1	2	3	4	5
21.	I have the ability to plan activities for students to author multimedia projects in Reading (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage).	1	2	3	4	5
22.	I am capable of using drill and practice software to reach instructional goals in Reading.	1	2	3	4	5
23.	I have the ability to integrate the use of e-mail into relevant Reading activities.	1	2	3	4	5
24.	I have the ability to plan Reading lessons that allow students to use the World Wide Web to access relevant sources for curriculum related projects.	1	2	3	4	5

1	2	3	4	5
0 - 19% Confident	**20 - 39% Confident**	**40 - 59% Confident**	**60 - 79% Confident**	**80 - 100% Confident**

How confident am I that ...

25.	I am capable of providing instruction in Math based on my students' interests.	1	2	3	4	5
26.	I am capable of providing Math instruction using a variety of hands-on manipulatives.	1	2	3	4	5
27.	I have the ability to adjust my classroom activities in Math to the learning needs of individual students.	1	2	3	4	5
28.	I am capable of providing appropriate learning experiences in Math for diverse learners (e.g., learning disabled, attention deficit, non-English speaking, gifted).	1	2	3	4	5
29.	I have the ability to adjust my classroom activities in Math to account for differences in individual students' skills.	1	2	3	4	5
30.	I am capable of providing timed practice tests for students during Math instruction.	1	2	3	4	5
31.	I have the ability to integrate Math activities into other curriculum areas.	1	2	3	4	5
32.	I have the ability to provide students with challenging seatwork in Math.	1	2	3	4	5
33.	I have the ability to prepare engaging materials to teach Math.	1	2	3	4	5
34.	I have the ability to plan effective lessons in Math.	1	2	3	4	5
35.	I have the ability to apply cooperative group learning strategies in teaching Math.	1	2	3	4	5
36.	I have the ability to provide worked examples during Math instruction.	1	2	3	4	5
37.	I am capable of conducting whole group instruction in Math.	1	2	3	4	5
38.	I am capable of conducting small group instruction in Math.	1	2	3	4	5
39.	I am capable of organizing learning activities in Math according to students' backgrounds.	1	2	3	4	5
40.	I am capable of selecting well-designed computer software to integrate into my Math curriculum.	1	2	3	4	5

1	2	3	4	5
0 - 19% Confident	20 - 39% Confident	40 - 59% Confident	60 - 79% Confident	80 - 100% Confident

How confident am I that ...

41. I am capable of providing learning activities for students that allow them to use computers or other technology to attain an instructional goal in Math. 1 2 3 4 5

42. I am capable of using computer tutorials to provide individualized Math instruction for students. 1 2 3 4 5

43. I have the ability to use computer games that match the goals of my Math curriculum. 1 2 3 4 5

44. I have the ability to plan activities for students to author multimedia projects in Math (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage). 1 2 3 4 5

45. I am capable of using drill and practice software to reach instructional goals in Math. 1 2 3 4 5

46. I have the ability to incorporate spreadsheet activities into my Math instruction. 1 2 3 4 5

47. I am capable of planning lessons that use videodiscs to reach an instructional goal in Math. 1 2 3 4 5

48. I have the ability to plan lessons that allow students to use the World Wide Web to access relevant sources for curriculum related projects in Math. 1 2 3 4 5

Part II Directions: Teachers are often more or less confident that successful implementation of specific instructional methods or strategies are effective in reaching instructional goals they have for their students. Using the scale provided, indicate your degree of confidence that the specific instructional method or strategy described for Reading or Math will be **effective in improving the standardized reading or math achievement scores** of the students you are presently teaching. For each item, circle the number that best represents your degree of confidence.

1	2	3	4	5
0 - 19% Confident	20 - 39% Confident	40 - 59% Confident	60 - 79% Confident	80 - 100% Confident

When implemented appropriately, how confident are you that an effective strategy for improving students' standardized reading achievement scores is ...

1. providing instruction based on students' interests 1 2 3 4 5

1	2	3	4	5
0 - 19% Confident	**20 - 39% Confident**	**40 - 59% Confident**	**60 - 79% Confident**	**80 - 100% Confident**

When implemented appropriately, how confident are you that an effective strategy for improving students' standardized reading achievement scores is ...

2.	integrating writing activities into the Reading instruction.	1	2	3	4	5
3.	adjusting classroom activities to the learning needs of individual students.	1	2	3	4	5
4.	providing appropriate learning experiences for diverse learners (e.g., learning disabled, attention deficit, non-English speaking, gifted).	1	2	3	4	5
5.	adjusting classroom activities to account for differences in individual students' Reading skills.	1	2	3	4	5
6.	providing instruction in phonics skills.	1	2	3	4	5
7.	integrating Reading activities into other curriculum areas.	1	2	3	4	5
8.	providing students with challenging seatwork.	1	2	3	4	5
9.	preparing engaging materials to teach Reading.	1	2	3	4	5
10.	planning effective lessons in Reading.	1	2	3	4	5
11.	applying cooperative group learning strategies.	1	2	3	4	5
12.	providing students with the opportunity to choose their own literature books for reading.	1	2	3	4	5
13.	conducting whole group instruction.	1	2	3	4	5
14.	conducting small group instruction.	1	2	3	4	5
15.	organizing learning activities according to students' backgrounds.	1	2	3	4	5
16.	integrating well-designed computer software into the Reading instruction.	1	2	3	4	5
17.	providing learning activities for students that allow them to use computers or other technology during instruction.	1	2	3	4	5
18.	using computer tutorials to provide individualized instruction for students in Reading.	1	2	3	4	5
19.	using CD-ROMs of children's literature that match the goals of the Reading curriculum.	1	2	3	4	5

1	2	3	4	5
0 - 19% Confident	20 - 39% Confident	40 - 59% Confident	60 - 79% Confident	80 - 100% Confident

When implemented appropriately, how confident are you that an effective strategy for improving students' standardized reading achievement scores is ...

20.	incorporating word processing activities (e.g., journal or story writing) into the instruction.	1	2	3	4	5
21.	planning activities for students to author multimedia projects in Reading (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage).	1	2	3	4	5
22.	using drill and practice software to reach instructional goals in Reading.	1	2	3	4	5
23.	integrating the use of e-mail into relevant Reading activities.	1	2	3	4	5
24.	planning lessons that allow students to use the World Wide Web to access relevant sources for curriculum related projects.	1	2	3	4	5

When implemented appropriately, how confident are you that an effective strategy for improving students' standardized math achievement scores is ...

25.	providing instruction based on my students' interests.	1	2	3	4	5
26.	providing instruction using a variety of hands-on manipulatives.	1	2	3	4	5
27.	adjusting classroom activities in Math to the learning needs of individual students.	1	2	3	4	5
28.	providing appropriate learning experiences for diverse learners (e.g., learning disabled, attention deficit, non-English speaking, gifted).	1	2	3	4	5
29.	adjusting classroom activities to account for differences in individual students' Math skills.	1	2	3	4	5
30.	providing timed practice tests for students during instruction.	1	2	3	4	5
31.	integrating Math activities into other curriculum areas.	1	2	3	4	5
32.	providing students with challenging seatwork in Math.	1	2	3	4	5
33.	preparing engaging materials to teach Math.	1	2	3	4	5
34.	planning effective lessons in Math.	1	2	3	4	5
35.	applying cooperative group learning strategies.	1	2	3	4	5

1	2	3	4	5
0 - 19% Confident	**20 - 39% Confident**	**40 - 59% Confident**	**60 - 79% Confident**	**80 - 100% Confident**

When implemented appropriately, how confident are you that an effective strategy for improving students' standardized math achievement scores is ...

36.	providing worked examples during instruction.	1	2	3	4	5
37.	conducting whole group instruction.	1	2	3	4	5
38.	conducting small group instruction.	1	2	3	4	5
39.	organizing learning activities according to students' backgrounds.	1	2	3	4	5
40.	integrating well-designed computer software into the Math curriculum.	1	2	3	4	5
41.	providing learning activities for students that allow them to use computers or other technology to attain an instructional goal in Math.	1	2	3	4	5
42.	using computer tutorials to provide individualized Math instruction for students.	1	2	3	4	5
43.	using computer games that match the goals of the Math curriculum.	1	2	3	4	5
44.	planning activities for students to author multimedia projects in Math (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage).	1	2	3	4	5
45.	using drill and practice software to reach instructional goals in Math.	1	2	3	4	5
46.	incorporating spreadsheet activities into Math instruction.	1	2	3	4	5
47.	planning lessons that use videodiscs to reach an instructional goal in Math.	1	2	3	4	5
48.	planning lessons that allow students to use the World Wide Web to access relevant sources for curriculum related projects in Math.	1	2	3	4	5

COMMENTS ABOUT THE INSTRUMENT

Were any of the directions unclear? If yes, which ones and do you have any suggestions?

Were any of the items unclear? If yes, which ones, why, and do you have any suggestions?

Any other comments or suggestions?

Appendix G

Survey Cover Letter

Survey Cover Letter

Dear Teacher,

You have been selected to participate in research examining teachers' beliefs about their ability to use a variety of instructional techniques during ____ instruction. This dissertation study is being conducted by Leslie Curda of the Department of Educational Psychology at the University of Oklahoma. If you choose to participate in this study I ask that you complete the enclosed questionnaire. The questionnaire asks you to provide a variety of demographic information about yourself and the students in your classroom. The survey also seeks your opinion about the usefulness of a variety of instructional strategies for teaching _____ and the confidence you have in your ability to perform these tasks successfully. It should take you approximately 15 minutes to complete the survey. At no time during the survey are you asked to reveal your name, your school name, or any of your students' names. Your responses to the questionnaire will be completely confidential.

The enclosed survey is being randomly distributed to elementary school teachers across the state of Tennessee. Should you choose to participate, please complete the enclosed questionnaire and return it using the self-addressed, stamped envelope provided. Returning the completed survey indicates your consent to participate in this study. Your participation in this study is voluntary. No negative consequences will result to those who decide not to participate. If at any time you change your mind about participating once you have begun, you may withdraw by simply choosing not to return the completed survey. Thank you for your time and your consideration.

Sincerely,

Leslie Curda

Appendix H

Reading Efficacy and Outcome Expectations Instrument

Reading Efficacy and Outcome Expectations Instrument

PART I: Please circle the appropriate answer or fill in the blank for each question below.

ABOUT YOURSELF:

GENDER: F M **AGE**: ________

RACE: African-American American Indian Asian-American Caucasian Hispanic

Other ____________________

EDUCATIONAL BACKGROUND: (Circle the highest one you have completed)

Bachelor's Degree Bachelor's + 15 Master's Degree Master's + 15 Doctorate

TEACHING EXPERIENCE:

How many years have you been teaching? ___

How many years have you been teaching the grade you are presently teaching? ___

ABOUT YOUR CLASSROOM AND STUDENTS:

What grade level are you presently teaching? K 1 2 3 4 5 6
Other (please specify) ____________

Do you teach both math and reading in the grade level you are presently teaching? YES NO

What approach do you use to teach reading? PHONICS WHOLE LANGUAGE OTHER______
(please specify)

How many students do you have in your present classroom? _____

For the following demographics indicate (estimate if necessary) **how many** of your students are in each category.

Gender: Male ___ Female ______

Race: African-American ______ American Indian ______ Asian-American ______

Caucasian ______ Other ______

SES: Free Lunch ______ Reduced Lunch ______

Reading: High Ability ___ Average Ability ______ Low Ability ___

Math: High Ability ___ Average Ability _____ Low Ability ___

How many of your students have specific individual learning needs in **math** identified through school testing services? _______

How many of your students have specific individual learning needs in **reading** identified through school testing services? _______

Students often vary in the way they participate and engage in activities in their class. In your opinion, approximately what **percentage** of the students in your class are actively engaged in the **math** lessons on a typical day? ______%

In your opinion, approximately what **percentage** of the students in your class are actively engaged in the **reading** lessons on a typical day? _______%

PART II: Directions: Circle the correct answer, fill in the blank, or use the scale provided to answer the following questions as accurately as possible.

1. Please rank order your preference for teaching the following subjects to your students. The subject you **most prefer** to teach should receive a rank of "1". Your next subject preference should receive a rank of "2" and so on up to a rank of "4".

   Science ___ Math ___ Reading ___ Social Studies ___

2. If you could choose to enroll in one of the following two courses during inservice teacher training, which one would you most prefer to attend? (circle one)

   A. A course which focuses on providing new ideas and teaching strategies for using the textbook and supporting materials in the subject area of your choice.

   B. A course which focuses on providing new ideas and strategies for integrating technology into the subject area of your choice.

3. Approximately how much time (in minutes) do you spend in a typical week **planning and preparing materials** for teaching Reading? _______ minutes

4. Approximately how much time (in minutes) do you spend in a typical week **engaged with your students in activities** related to Reading? _______ minutes

5. Approximately how much time (in minutes) do you spend in a typical week **planning and preparing to integrate the use of technology** into your curriculum? _______ minutes

6. Approximately how much time (in minutes) do you spend in a typical **week engaged with your students in activities related to technology** use? _______ minutes

Please respond to the following statements using the scale provided.

1	2	3	4	5	6
Strongly Disagree	**Disagree**	**Somewhat Disagree**	**Somewhat Agree**	**Agree**	**Strongly Agree**

		SD	D	SWD	SWA	A	SA
7.	Compared to the normal effort I exert teaching Reading, I put more effort into working with a student who is not making progress in his/her reading ability.	1	2	3	4	5	6
8.	Compared to the normal effort I exert teaching Reading, I put more effort into developing enrichment activities for students who have above average reading ability.	1	2	3	4	5	6
9.	Compared to the normal effort I exert planning in other subject areas, I put more effort into planning daily Reading instruction.	1	2	3	4	5	6
10.	I continue working with a student until he or she learns essential reading skills.	1	2	3	4	5	6
11.	I continue to try a variety of instructional strategies until I find a strategy that is effective with my students in Reading.	1	2	3	4	5	6
12.	I try a variety of approaches to teaching Reading in order to discover strategies that motivate my students to improve their reading ability.	1	2	3	4	5	6
13.	I feel I have control over my students' achievement in Reading.	1	2	3	4	5	6
14.	Teaching Reading is stressful.	1	2	3	4	5	6
15.	I think teachers in my building can affect student achievement in Reading more than me.	1	2	3	4	5	6
16.	I feel responsible for my students' achievement in Reading.	1	2	3	4	5	6
17.	Compared to the normal effort I exert planning in other subject areas, I put more effort into trying to integrate technology into my classroom curriculum.	1	2	3	4	5	6
18.	Compared to the normal effort I exert preparing for instruction in other subject areas, I put more effort into identifying specific computer applications that will enhance student learning experiences.	1	2	3	4	5	6
19.	Compared to the normal effort I exert developing instructional activities using the textbook, I put more effort into developing instructional activities that allow students to use technology to achieve a curriculum goal.	1	2	3	4	5	6

	SD	D	SWD	SWA	A	SA
20. I continue to try integrating technology into my curriculum even when I am faced with computer mishaps or equipment problems.	1	2	3	4	5	6
21. I continue to try new strategies for integrating technology into my curriculum even though I have a lack of experience with technology.	1	2	3	4	5	6
22. I continue to try approaches to integrating technology even when I feel like I experience more classroom management problems.	1	2	3	4	5	6

Part III Directions: Teachers are often more or less confident they can successfully employ specific instructional methods or strategies with their students towards the goal of improving their achievement in Reading. Using the scale provided, indicate your level of agreement for successfully using the instructional methods or strategies described with the students you are presently teaching. As with many teachers, you may not actively employ some of the strategies listed below. For example, you may be using a whole language approach to teach reading or you may not have access to certain types of technology; however, you should still be able to report your confidence that you could successfully use a basal reader or apply the technology if you had access. Remember that the focus of these questions is your confidence in using the instructional strategies. These questions are not asking if you are presently using these strategies or agree with these strategies. For each item, circle the number that best represents your level of agreement to each statement.

1	2	3	4	5	6
Strongly Disagree	**Disagree**	**Somewhat Disagree**	**Somewhat Agree**	**Agree**	**Strongly Agree**

I am confident that...

	SD	D	SWD	SWA	A	SA
1. I have the ability to apply cooperative group learning strategies in teaching Reading.	1	2	3	4	5	6
2. I am capable of integrating writing activities into my Reading instruction.	1	2	3	4	5	6
3. I am capable of conducting whole group instruction in Reading.	1	2	3	4	5	6
4. I have the ability to integrate Reading activities into other curriculum areas.	1	2	3	4	5	6
5. I am capable of providing instruction in phonics skills when teaching Reading.	1	2	3	4	5	6
6. I have the ability to plan effective lessons in Reading.	1	2	3	4	5	6
7. I am capable of using drill and practice software to reach instructional goals in Reading.	1	2	3	4	5	6
8. I am capable of providing instruction in Reading based on students' interests.	1	2	3	4	5	6

I am confident that...

	SD	D	SWD	SWA	A	SA
9. I am capable of conducting small group instruction in Reading.	1	2	3	4	5	6
10. I am capable of providing students with the opportunity to choose their own literature books for Reading instruction.	1	2	3	4	5	6
11. I have the ability to prepare engaging materials to teach Reading.	1	2	3	4	5	6
12. I have the ability to incorporate word processing activities (e.g., journal or story writing) into my Reading instruction.	1	2	3	4	5	6
13. I have the ability to provide students with challenging seatwork in Reading.	1	2	3	4	5	6
14. I am capable of providing appropriate learning experiences in Reading for diverse learners (e.g., learning disabled, attention deficit, non-English speaking, gifted).	1	2	3	4	5	6
15. I am capable of selecting well-designed computer software to integrate into Reading instruction.	1	2	3	4	5	6
16. I am capable of using computer tutorials to provide individualized instruction in Reading.	1	2	3	4	5	6
17. I am capable of providing learning activities for students that allow them to use computers or other technology during Reading instruction.	1	2	3	4	5	6
18. I have the ability to adjust my classroom activities in Reading to account for differences in individual students' skills.	1	2	3	4	5	6
19. I have the ability to adjust my classroom activities in Reading to the learning needs of individual students.	1	2	3	4	5	6
20. I am capable of using CD-ROMs that contain children's literature books that match the goals of my Reading curriculum.	1	2	3	4	5	6
21. I have the ability to integrate the use of e-mail into relevant Reading activities.	1	2	3	4	5	6
22. I am capable of organizing learning activities in Reading according to students' backgrounds.	1	2	3	4	5	6
23. I have the ability to plan Reading lessons that allow students to use the World Wide Web to access relevant sources for curriculum related projects.	1	2	3	4	5	6
24. I have the ability to plan activities for students to author multimedia projects in Reading (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage).	1	2	3	4	5	6

Part IV Directions: Teachers are often more or less confident that successful implementation of specific instructional methods or strategies are effective in reaching instructional goals they have for their students. Using the scale provided, indicate your level of agreement that the specific instructional method or strategy described for Reading will be **effective in improving the standardized reading achievement scores** of the students you are presently teaching. For each item, circle the number that best represents your level of agreement.

1	2	3	4	5	6
Strongly Disagree	**Disagree**	**Somewhat Disagree**	**Somewhat Agree**	**Agree**	**Strongly Agree**

I am confident that an effective strategy for improving my students' standardized reading achievement scores is...

	SD	D	SWD	SWA	A	SA
1. applying cooperative group learning strategies in Reading.	1	2	3	4	5	6
2. integrating writing activities into the Reading instruction.	1	2	3	4	5	6
3. conducting whole group instruction in Reading.	1	2	3	4	5	6
4. integrating Reading activities into other curriculum areas.	1	2	3	4	5	6
5. providing instruction in phonics skills.	1	2	3	4	5	6
6. planning effective lessons in Reading.	1	2	3	4	5	6
7. using drill and practice software to reach instructional goals in Reading.	1	2	3	4	5	6
8. providing Reading instruction based on students' interests.	1	2	3	4	5	6
9. conducting small group instruction in Reading.	1	2	3	4	5	6
10. providing students with the opportunity to choose their own literature books for Reading.	1	2	3	4	5	6
11. preparing engaging materials to teach Reading.	1	2	3	4	5	6
12. incorporating word processing activities (e.g., journal or story writing) into the Reading instruction.	1	2	3	4	5	6
13. providing students with challenging seatwork in Reading.	1	2	3	4	5	6
14. providing appropriate learning experiences for diverse learners (e.g., learning disabled, attention deficit, non-English speaking, gifted).	1	2	3	4	5	6
15. integrating well-designed computer software into Reading instruction.	1	2	3	4	5	6

I am confident that an effective strategy for improving my students' standardized reading achievement scores is...

	SD	D	SWD	SWA	A	SA
16. using computer tutorials to provide individualized instruction for students in Reading.	1	2	3	4	5	6
17. providing learning activities for students that allow them to use computers or other technology during Reading instruction.	1	2	3	4	5	6
18. adjusting classroom activities to account for differences in individual students' Reading skills.	1	2	3	4	5	6
19. adjusting classroom Reading activities to the learning needs of individual students.	1	2	3	4	5	6
20. using CD-ROMs of children's literature that match the goals of the Reading curriculum.	1	2	3	4	5	6
21. integrating the use of e-mail into relevant Reading activities.	1	2	3	4	5	6
22. organizing learning activities according to students' backgrounds.	1	2	3	4	5	6
23. planning lessons that allow students to use the World Wide Web to access relevant sources for Reading related projects.	1	2	3	4	5	6
24. planning activities for students to author multimedia projects in Reading (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage).	1	2	3	4	5	6

Appendix I

Math Efficacy and Outcome Expectations Instrument

Math Efficacy and Outcome Expectations Instrument

PART I: Please circle the appropriate answer or fill in the blank for each question below.

ABOUT YOURSELF:

GENDER: F M **AGE:** _______

RACE: African-American American Indian Asian-American Caucasian Hispanic

Other ___________________

EDUCATIONAL BACKGROUND: (Circle the highest one you have completed)

Bachelor's Degree Bachelor's + 15 Master's Degree Master's + 15 Doctorate

TEACHING EXPERIENCE:

How many years have you been teaching? ___

How many years have you been teaching the grade you are presently teaching? ___

ABOUT YOUR CLASSROOM AND STUDENTS:

What grade level are you presently teaching? K 1 2 3 4 5 6
Other (please specify) ____________

Do you teach both math and reading in the grade level you are presently teaching? YES NO

What approach do you use to teach mathematics? _____________________

How many students do you have in your present classroom? ______

For the following demographics indicate (estimate if necessary) **how many** of your students are in each category.

Gender: Male ___ Female ______

Race: African-American _____ American Indian _____ Asian-American _____

Caucasian _____ Other ______

SES: Free Lunch _____ Reduced Lunch _____

Reading: High Ability ___ Average Ability _____ Low Ability ___

Math: High Ability ___ Average Ability _____ Low Ability ___

How many of your students have specific individual learning needs in **math** identified through school testing services? _______

How many of your students have specific individual learning needs in **reading** identified through school testing services? _______

Students often vary in the way they participate and engage in activities in their class. In your opinion, approximately what **percentage** of the students in your class are actively engaged in the **math** lessons on a typical day? ______%

In your opinion, approximately what **percentage** of the students in your class are actively engaged in the **reading** lessons on a typical day? _______%

PART II: Directions: Circle the correct answer, fill in the blank, or use the scale provided to answer the following questions as accurately as possible.

1. Please rank order your preference for teaching the following subjects to your students. The subject you **most prefer** to teach should receive a rank of "1". Your next subject preference should receive a rank of "2" and so on up to a rank of "4".

   Science ___ Math ___ Reading ___ Social Studies ___

2. If you could choose to enroll in one of the following two courses during inservice teacher training, which one would you most prefer to attend? (circle one)

   A. A course which focuses on providing new ideas and teaching strategies for using the textbook and supporting materials in the subject area of your choice.

   B. A course which focuses on providing new ideas and strategies for integrating technology into the subject area of your choice.

3. Approximately how much time (in minutes) do you spend in a typical week **planning and preparing materials** for teaching Math? _______ minutes

4. Approximately how much time (in minutes) do you spend in a typical week **engaged with your students in activities** related to Math? _______ minutes

5. Approximately how much time (in minutes) do you spend in a typical week **planning and preparing to integrate the use of technology** into your curriculum? _______ minutes

6. Approximately how much time (in minutes) do you spend in a typical **week engaged with your students in activities related to technology** use? _______ minutes

Please respond to the following statements using the scale provided.

1	2	3	4	5	6
Strongly Disagree	**Disagree**	**Somewhat Disagree**	**Somewhat Agree**	**Agree**	**Strongly Agree**

	SD	D	SWD	SWA	A	SA
7. Compared to the normal effort I exert teaching Math, I put more effort into working with a student who is having trouble grasping a new math concept or procedure.	1	2	3	4	5	6
8. Compared to the normal effort I exert teaching Math, I put more effort into developing enrichment activities for students who have above average math ability.	1	2	3	4	5	6
9. Compared to the normal effort I exert planning in other subject areas, I put more effort into planning daily Math instruction.	1	2	3	4	5	6
10. I continue working with a student until he or she learns essential math skills.	1	2	3	4	5	6
11. I continue to try a variety of instructional strategies until I find a strategy that is effective with my students in Math.	1	2	3	4	5	6
12. I try a variety of approaches to teaching Math in order to discover strategies that motivate my students to improve their understanding of Math.	1	2	3	4	5	6
13. I think teachers in my building can affect student achievement in Math more than me.	1	2	3	4	5	6
14. I feel responsible for my students' achievement in Math.	1	2	3	4	5	6
15. Teaching Math is stressful.	1	2	3	4	5	6
16. I feel I have control over my students' achievement in Math.	1	2	3	4	5	6
17. Compared to the normal effort I exert planning in other subject areas, I put more effort into trying to integrate technology into my classroom curriculum.	1	2	3	4	5	6
18. Compared to the normal effort I exert preparing for instruction in other subject areas, I put more effort into identifying specific computer applications that will enhance student learning experiences.	1	2	3	4	5	6
19. Compared to the normal effort I exert developing instructional activities using the textbook, I put more effort into developing instructional activities that allow students to use technology to achieve a curriculum goal.	1	2	3	4	5	6

	SD	D	SWD	SWA	A	SA
20. I continue to try integrating technology into my curriculum even when I am faced with computer mishaps or equipment problems.	1	2	3	4	5	6
21. I continue to try new strategies for integrating technology into my curriculum even though I have a lack of experience with technology.	1	2	3	4	5	6
22. I continue to try approaches to integrating technology even when I feel like I experience more classroom management problems.	1	2	3	4	5	6

Part III Directions: Teachers are often more or less confident they can successfully employ specific instructional methods or strategies with their students towards the goal of improving their achievement in Math. Using the scale provided, indicate your level of agreement for successfully using the instructional methods or strategies described with the students you are presently teaching. As with many teachers, you may not actively employ some of the strategies listed below. For example, you may be using a whole language approach to teach reading or you may not have access to certain types of technology; however, you should still be able to report your confidence that you could successfully use a basal reader or apply the technology if you had access. Remember that the focus of these questions is your confidence in using the instructional strategies. These questions are not asking if you are presently using these strategies or agree with these strategies. For each item, circle the number that best represents your level of agreement to each statement.

1	2	3	4	5	6
Strongly Disagree	**Disagree**	**Somewhat Disagree**	**Somewhat Agree**	**Agree**	**Strongly Agree**

I am confident that...

	SD	D	SWD	SWA	A	SA
1. I have the ability to provide worked examples of problems during Math instruction.	1	2	3	4	5	6
2. I am capable of conducting whole group instruction in Math.	1	2	3	4	5	6
3. I have the ability to apply cooperative group learning strategies in teaching Math.	1	2	3	4	5	6
4. I am capable of providing Math instruction using a variety of hands-on manipulatives.	1	2	3	4	5	6
5. I am capable of providing timed practice tests for students during Math instruction.	1	2	3	4	5	6
6. I am capable of using drill and practice software to reach instructional goals in Math.	1	2	3	4	5	6
7. I have the ability to integrate Math activities into other curriculum areas.	1	2	3	4	5	6
8. I have the ability to plan effective lessons in Math.	1	2	3	4	5	6

I am confident that...

	SD	D	SWD	SWA	A	SA
9. I have the ability to provide students with challenging seatwork in Math.	1	2	3	4	5	6
10. I am capable of conducting small group instruction in Math.	1	2	3	4	5	6
11. I have the ability to prepare engaging materials to teach Math.	1	2	3	4	5	6
12. I am capable of providing learning activities for students that allow them to use computers or other technology to attain an instructional goal in Math.	1	2	3	4	5	6
13. I have the skills needed to adjust my classroom activities in Math to the learning needs of individual students.	1	2	3	4	5	6
14. I have the ability to use computer games that match the goals of my Math curriculum.	1	2	3	4	5	6
15. I have the ability to adjust my classroom activities in Math to account for differences in individual students' skills.	1	2	3	4	5	6
16. I am capable of using computer tutorials to provide individualized Math instruction for students.	1	2	3	4	5	6
17. I am capable of selecting well-designed computer software to integrate into my Math curriculum.	1	2	3	4	5	6
18. I am capable of providing appropriate learning experiences in Math for diverse learners (e.g., learning disabled, attention deficit, non-English speaking, gifted).	1	2	3	4	5	6
19. I have the skills to provide instruction in Math based on my students' interests.	1	2	3	4	5	6
20. I am capable of organizing learning activities in Math according to students' backgrounds.	1	2	3	4	5	6
21. I have the ability to incorporate spreadsheet activities into my Math instruction.	1	2	3	4	5	6
22. I have the skills to plan Math lessons that allow students to use the World Wide Web to access relevant sources for curriculum related projects.	1	2	3	4	5	6
23. I am capable of planning lessons that use videodiscs to reach an instructional goal in Math.	1	2	3	4	5	6
24. I have the ability to plan activities for students to author multimedia projects in Math (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage).	1	2	3	4	5	6

Part IV Directions: Teachers are often more or less confident that successful implementation of specific instructional methods or strategies are effective in reaching instructional goals they have for their students. Using the scale provided, indicate your level of agreement that the specific instructional method or strategy described for Math will be **effective in improving the standardized math achievement scores** of the students you are presently teaching. For each item, circle the number that best represents your level of agreement.

1	2	3	4	5	6
Strongly Disagree	**Disagree**	**Somewhat Disagree**	**Somewhat Agree**	**Agree**	**Strongly Agree**

I am confident that an effective strategy for improving my students' standardized math achievement scores is...

		SD	D	SWD	SWA	A	SA
1.	providing worked examples of problems during Math instruction.	1	2	3	4	5	6
2.	conducting whole group instruction in Math.	1	2	3	4	5	6
3.	applying cooperative group learning strategies in Math.	1	2	3	4	5	6
4.	providing Math instruction using a variety of hands-on manipulatives.	1	2	3	4	5	6
5.	providing timed practice tests for students during Math instruction.	1	2	3	4	5	6
6.	using drill and practice software to reach instructional goals in Math.	1	2	3	4	5	6
7.	planning effective lessons in Math.	1	2	3	4	5	6
8.	integrating Math activities into other curriculum areas.	1	2	3	4	5	6
9.	providing students with challenging seatwork in Math.	1	2	3	4	5	6
10.	conducting small group instruction in Math.	1	2	3	4	5	6
11.	preparing engaging materials to teach Math.	1	2	3	4	5	6
12.	providing learning activities for students that allow them to use computers or other technology to attain an instructional goal in Math.	1	2	3	4	5	6
13.	adjusting classroom Math activities to the learning needs of individual students.	1	2	3	4	5	6
14.	using computer games that match the goals of the Math curriculum.	1	2	3	4	5	6
15.	adjusting classroom activities to account for differences in individual students' Math skills.	1	2	3	4	5	6

I am confident that an effective strategy for improving my students' standardized math achievement scores is...

	SD	D	SWD	SWA	A	SA
16. using computer tutorials to provide individualized Math instruction for students.	1	2	3	4	5	6
17. integrating well-designed computer software into Math instruction.	1	2	3	4	5	6
18. providing appropriate Math learning experiences for diverse learners (e.g., learning disabled, attention deficit, non-English speaking, gifted).	1	2	3	4	5	6
19. providing Math instruction based on my students' interests.	1	2	3	4	5	6
20. organizing learning activities in Math according to students' backgrounds.	1	2	3	4	5	6
21. incorporating spreadsheet activities into Math instruction.	1	2	3	4	5	6
22. planning lessons that allow students to use the World Wide Web to access relevant sources for Math related projects.	1	2	3	4	5	6
23. planning lessons that use videodiscs to reach an instructional goal in Math.	1	2	3	4	5	6
24. planning activities for students to author multimedia projects in Math (e.g., using HyperCard or HyperStudio or building a World Wide Web HomePage).	1	2	3	4	5	6