DESIGN AND VALIDATION OF A STANDARDS-BASED SCIENCE TEACHER EFFICACY INSTRUMENT

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree of Doctor of Philosophy in the Graduate

School of The Ohio State University

By

Patricia Reda Kerr, B.S., M.S.

* * * * * * * * * *

The Ohio State University
2006

Dissertation Committee:

Professor Donna F. Berlin, Advisor

Professor Arthur L. White

Professor Garry McKenzie

Approved by

____

Advisor
College of Education

ABSTRACT

National standards for K-12 science education address all aspects of science education, with their main emphasis on curriculum--both science subject matter and the process involved in doing science. Standards for science teacher education programs have been developing along a parallel plane, as is self-efficacy research involving classroom teachers. Generally, studies about efficacy have been dichotomous--basing the theoretical underpinnings on the work of either Rotter's Locus of Control theory or on Bandura's explanations of efficacy beliefs and outcome expectancy.

This study brings all three threads together--K-12 science standards, teacher education standards, and efficacy beliefs--in an instrument designed to measure science teacher efficacy with items based on identified critical attributes of standards-based science teaching and learning. Based on Bandura's explanation of efficacy being task-specific and having outcome expectancy, a developmental, systematic progression from standards-based strategies and activities to tasks to critical attributes was used to craft items for a standards-based science teacher efficacy instrument. Demographic questions related to school characteristics, teacher characteristics, preservice background, science teaching experience, and post-certification professional development were included in the instrument. The instrument was completed by 102 middle level science teachers, with complete data for 87 teachers.

A principal components analysis of the science teachers' responses to the instrument resulted in two components: Standards-Based Science Teacher Efficacy: Beliefs About Teaching (BAT, reliability = .92) and Standards-Based Science Teacher Efficacy: Beliefs About Student

Achievement (BASA, reliability = .82). Variables that were characteristic of professional development activities, science content preparation, and school environment were identified as members of the sets of variables predicting the BAT and BASA subscales. Correlations were computed for BAT, BASA, and demographic variables to identify relationships between teacher efficacy, teacher characteristics, and school characteristics.

Further research is recommended to refine the instrument and apply its use to a larger sample of science teachers. Its further development also has significance for the enhancement of science teacher education programs.

Dedicated to Jamey and Micha

ACKNOWLEDGMENTS

I thank my advisor, Donna Berlin, for all her time, encouragement, and editing skills. As a true mentor, her support exceeded the academic mandates, and her compassion strengthened my resolve.

I thank my committee member, Art White, for his patience in shepherding me through the statistics.

I thank my committee member, Garry McKenzie, for continually helping me to recalibrate my perspective and who inspired my interest in Earth Science.

I also extend my gratitude to a group of women whom I may not name, but who will never be anonymous.

VITA

August 11, 1948 .... Born-San Antonio, Texas

1970 .... B.S. Elementary Education, Slippery Rock State College, Pennsylvania

1970-1990.. Classroom Teacher

1990 .... M.S. Science Education, Clarion State University, Pennsylvania

1990-2000.. K-12 Science Specialist, Kentucky Department of Education

2000-Present.. Graduate Teaching and Research Associate, College of Education, School of Teaching and Learning, The Ohio State University

FIELDS OF STUDY

Major Field: College of Education
Teaching and Learning: Science Education

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The last years of the Cold War signaled the beginning of a change in perceptions toward schools in general, and science in particular. It brought less interest in nurturing future scientists and mathematicians, and more interest in science and mathematics literacy for all students. Scores from standardized tests gained in significance both with heavy marketing from the test producing-companies and a higher level of accountability for student success in school. Interest in continuity and consistency of school curricula took on new importance as families became increasingly more mobile in search of a living wage. Standardization of curricula, instruction, and assessments moved to center stage, along with criticisms of teachers who were inadequately prepared.

The publication of national standards for mathematics, led by the National Council of Teachers of Mathematics (1989), was a turning point for science education. While their process was not without difficulties, the publication of standards for mathematics education was the result of a more-or-less unified community. The development of standards for science education was not so smooth, one reason being that the advocates for a scientifically-literate society did not agree about how the standards should be articulated. Among other issues, mathematics educators did not face the prospect of certification and/or licensure in separate disciplines of mathematics (algebra, geometry, and trigonometry) as science educators faced for the Life Sciences (biology), Physical Sciences (chemistry and physics separately), and the Earth and Space Sciences (mainly geology). The American Association for the Advancement of Science (AAAS) emerged as a dominant voice for national standards for science education. In 1989, the AAAS published *Science for All Americans.* It was the first publication in a long-term initiative to advance scientific, mathematical, and technological literacy. In this document, AAAS set down a series of recommendations on scientific understandings and ways of thinking that are essential for all modern citizens. In 1993, AAAS completed and published its next installment -- *Benchmarks for Science Literacy*. *Benchmarks* became the foundation upon which later standards documents from the National Academy of Sciences (NAS) and the National Research Council (NRC) grew. The NRC publication *National Science Education Standards* (National Research Council, 1996), or NSES, was essentially a guide for K-12 curriculum development, whereas *Benchmarks* described what a scientifically-literate society should know, understand, and be able to do. Although the organizational structure for each document was considerably different from the other, and different topics were addressed in each, the discipline-specific knowledge expectations for students in grades K-12 had an overlap of more than 90% (Nelson, 1997).

Rationale

A teacher's belief in herself/himself as a professional is a fundamental factor in the schema of her/his worldview. Since the effects of these self-efficacy beliefs ripple outward to an ever-larger boundary of influence, it is essential to be aware of situations and experiences that shape self-efficacy beliefs. For example, if a teacher preparation program does not offer a variety of rich field experiences, the preservice teacher's beliefs about a classroom learning environment and beliefs about what students can accomplish are based on imagination, not actual experiences.

There is evidence that this disconnection leads to disillusionment and early departure from teaching. It also influences the types and frequency of professional activities in which a teacher will participate, among other things. Therefore, it is essential to shape self-efficacy beliefs with preservice learning experiences that are realistic, positive, and supportive. One of the ways in which this goal can be accomplished is to determine preservice and inservice teacher self-efficacy beliefs with an instrument that is aligned with the standards-based expectations of the profession and local curriculum.

The *National Science Education Standards* (NRC, 1996) are the context for this study. The standards have been in existence since the mid-1990s, and they are the basis for national, state, and local reform efforts. This publication continues to influence changes in curriculum design, instructional strategies, assessment and evaluation, and professional development. With a level of scrutiny directed at the preponderance of development affected by the NSES, it has also resulted in the identification of various gaps or delayed connections. For example, the evaluation and accreditation of science teacher preparation programs are also based on the NSES, but one set of program standards is undergoing its second revision since its initial development in 1998 (National Science Teachers Association, 2003). This hardly allows sufficient time for college and university programs that prepare science teachers to establish and refine their programs before another version of teacher education program standards appears. It also confuses the alignment of the K-12 expectations for science with the expectations of the teacher preparation programs and professional development providers. Eventually, questions arise about the relationships that exist among the components of curriculum, instruction, and assessment in K-12 education and teacher professional development, both before and after initial certification and/or licensure.

It is of interest to examine the relationship between standards for K-12 science education and science teacher preparation programs. It is also of interest to examine a connection between these two and a teacher's sense of self-efficacy. Studies show the significance of a strong sense of self-efficacy to teacher perseverance in the profession (Caprara, Barbaranelli, & Borgogni, 2003; Parker, Guarino, & Smith, 2002), to the likelihood of continuing to grow professionally (Roberts, Henson, & Tharp, 2001), and to connections between teacher beliefs about self-efficacy and beliefs about student achievement (Lawrenz, 1975). However, there does not seem to be literature that identifies or describes relationships that exist among standards-based science, science teacher preparation programs, teacher beliefs about their sense of self-efficacy, and their beliefs about student achievement in standards-based science. This is a complex relationship, and can be described through a series of questions.

1. Are state-level science assessments designed from science education standards?
2. If classroom science curriculum is aligned with standards-based science, and teachers are expected to use instructional strategies that are also aligned with science standards, do science teacher preparation programs adequately address the curriculum, instructional strategies, and assessments they are expected to use?
3. What is the level of importance that teacher practitioners attach to attributes of standards-based science?
4. How important do practitioners think it is to explicitly address attributes of standards-based science in their preparation programs?
5. Are teacher candidate experiences as learners in congruence with standards-based science, and how do their experiences affect their self-efficacy beliefs?
6. How are teacher self-efficacy beliefs, in the context of standards-based science, related to their beliefs about student achievement?

When reading the aforementioned questions, the progression appears to be linear, but it is more circular. The last question refers to student achievement in science. It connects back to the very thing that is measured in state assessments referred to in the first question. The progression moves as illustrated in Figure 1.1.

Interconnectedness of Standards, Teacher Self-Efficacy Beliefs and Student Achievement

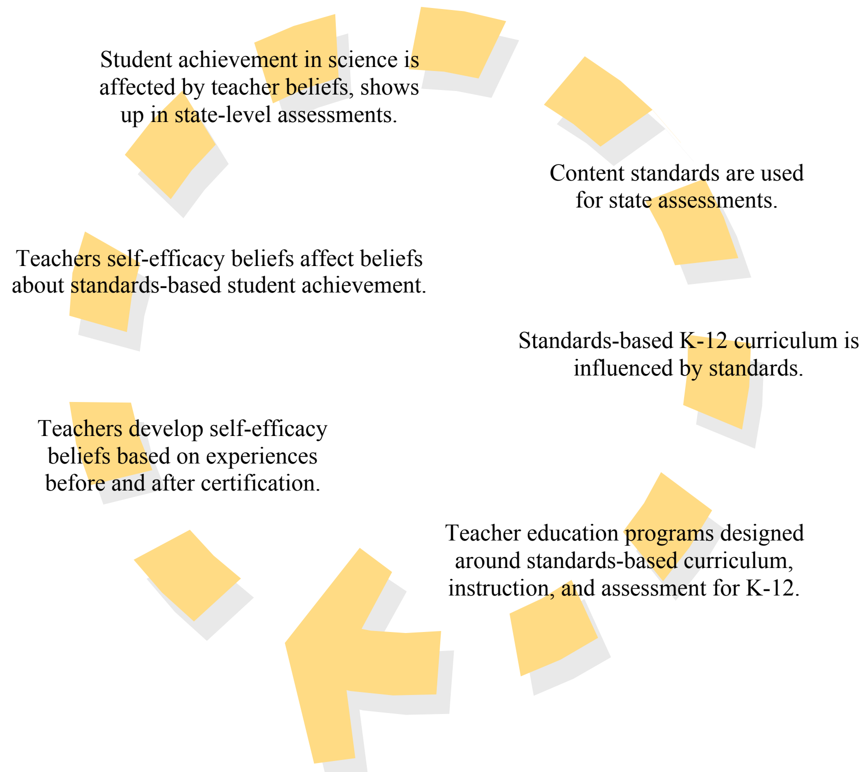


Figure 1.1: The interconnectedness of standards, teacher self-efficacy beliefs, and student achievement

An additional component of this recursive process is the importance of explicitly including attributes of standards-based science education in teacher preparation programs, rather than the belief that they are implicit in every course syllabus. The standards for accrediting college and university programs that prepare science teachers call for an assessment system to measure the knowledge (both content and pedagogical), skills, and dispositions of teacher candidates (National Science Teachers Association, 2003). Again, a missing piece is the one relating to a sense of self-efficacy and its connections to and influence on other components.

The Evolution of Teacher Self-Efficacy Instruments

Overview of Relevant Literature

Although self-efficacy has been the topic of many research studies since the mid-1960s, recently several studies have examined the evolution of self-efficacy-related research (Henson, 2002; Soodak & Podell, 1996; Tschannen-Moran & Woolfolk Hoy, 2001). Examining the changes in self-efficacy related research over the past three decades has resulted in an opportunity for a fresh look at original ideas and definitions. Recent findings in brain research and cognitive learning theory contributed to new insights, interpretations, and applications of the tenets of self-efficacy research. One challenge to self-efficacy research is an evolving realization of its complexity. A second consistent focus of research has been the development of valid, reliable instruments. A third aspect of the research is the attempt to determine the relationships between self-efficacy (and beliefs about self-efficacy) and the consequences and implications of those beliefs. If self-efficacy is as complex as it seems, designing instruments to measure it may be fragmented at best and determining relationships between self-efficacy and specific consequences is even more complicated.

The evolution of instruments moves along a timeline that began with Armor et al.'s Rand Study in 1976, which was based on Rotter's social learning theory of 1966. The Rand researchers had only two questions to which teachers responded. One referred to external control; the other to internal control. The first item, beliefs about external factors influencing teachers and schools, is labeled *general teaching efficacy* and the second item, beliefs about the efficacy of their own teaching, is labeled *personal teaching efficacy*. The sum of the two items was called *teacher efficacy*. In Rotter's theory, locus of control is either outside the teacher's control (external) or inside the teacher's control (internal) (Armor, et al.; Rotter).

Three subsequent instruments were all grounded in Rotter's social learning theory. Guskey developed a 30-item instrument measuring Responsibility for Student Achievement

(RSA), student success (which was labeled R+), and student failure (labeled R-). Rose and Medway (1981) developed a 28-item measure called Teacher Locus of Control (TLC.) Ashton (1982) expanded the Rand questions in a refinement of the *Webb Efficacy Scale.* Guskey's, Rose and Medway's, and Ashton's instruments came into a somewhat wider practice after the initial studies were completed, and self-efficacy research continued to search for clarification of the construct.

Additional clarification emerged from Albert Bandura's development of social learning theory in 1971, and the development of the construct of self-efficacy in 1977. Bandura defined perceived self-efficacy as "... beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (Bandura, 1971, p. 83)--a person's conviction that s/he has the necessary knowledge and skills to perform a specific task. Outcome expectancy emerged from the social cognitive theory and connected efficacy beliefs to an expectation of certain outcomes if the task is performed competently (Bandura, 1986).

Studies by Gibson and Dembo (1984) and also Ashton and Webb (1983) both attempted to incorporate the idea of internal/external control with Bandura's construct of self-efficacy. Ashton and Webb's earlier study (1986) yielded two dimensions: (a) *teaching efficacy* (Bandura's outcome expectations) and (b) *personal efficacy* (Bandura's efficacy expectation). Gibson and Dembo developed a 30-item scale titled the Teacher Efficacy Scale (TES) yielding a two-factor structure of previously identified dimensions, which they labeled *general teaching efficacy* (GTE) and *personal teaching efficacy* (PTE). They concluded that GTE was similar to Bandura's outcome expectations dimension, and PTE was similar to Bandura's efficacy expectations. Ashton, Olejnik, Crocker, and McAuliffe (1982) also developed vignettes to describe situations and asked teachers to make judgments about how they would perform considering two different frames of reference - themselves and other teachers. Gibson and Dembo's scale survived and appeared in various modified forms, while Ashton's vignettes fell into disuse.

Each iteration of the instruments based its design on a definition that was amorphous, and, over time, studies targeted increasingly more specific populations. Much valuable information, related to teacher satisfaction in the profession, student achievement, classroom environment, teacher longevity, and professional development, resulted from this research. Perhaps the progress of research on self-efficacy beliefs is analogous to research about atoms - scientists are not absolutely certain what one looks like, but they understand its fundamental behavior and can discuss relationships between the structure of the atom and situations that occur as a result of the atom's configuration.

Since Bandura's (1977) work on the development of the concept of self-efficacy in the mid-1970s, instruments designed to measure a teacher's sense of self-efficacy vary in task-specificity when measuring general teaching efficacy or measuring personal teaching efficacy. The Science Teaching Efficacy Beliefs Instrument (STEBI) by Enochs and Riggs (1990) is an example of an instrument with greater specificity that aligns with Bandura's recommendation for domain-specific instruments. It identifies subject, grade-level, and experience-level criteria to filter respondents. At this time, there seems to be a gap between efficacy belief instruments and the *National Science Education Standards* (NRC, 1996). Instruments related to science teaching are not incorporating the philosophical, pedagogical, or teaching environment issues raised by the standards. How can we measure teachers' efficaciousness in a standards-based teaching environment if no such instrument exists? A measurement component connecting self-efficacy to standards also seems to be missing from teacher preparation programs. Therefore, it was the intention of this study to develop an instrument to determine relationships, if any, between the following, in the context of the *National Science Education Standards* (NRC).

1. Science teacher perceptions about the level of importance of addressing certain components or attributes of the standards in teacher preparation programs.

2. Science teacher beliefs in their efficaciousness as science teachers, using the attributes of the standards.
3. The relationships between teacher beliefs in their self-efficacy and their beliefs about student achievement, relative to the identified attributes of the standards.

According to Bandura (1997) and Tschannen-Moran and Woolfolk Hoy (2001) as well as Henson (2002), self-efficacy beliefs develop within contexts. For example, a person may have a strong sense of efficacy about playing tennis in a friendly game against a neighbor, but have a very different sense of self-efficacy if the opponent is a Wimbledon champion. A teacher's sense of self-efficacy is related to context as well. For example, a teacher may have a strong sense of self-efficacy about teaching science at the primary grades, but a weak sense of self-efficacy about teaching science at the middle grades or at the high school level. The teacher may believe s/he has an adequate background of knowledge in a science discipline to teach concepts to beginning learners, but believe s/he is less efficacious when expected to teach concepts at a more sophisticated level with students who are more experienced in science.

Science teacher preparation programs began to evolve in alignment with the standards, but a gap developed between the structure of teacher preparation programs and what teachers face when they have their own classrooms. Several studies reveal new teachers' discomfort upon discovering that the confidence they developed during their preservice education begins to erode (Luft & Patterson, 2002; Raizen & Michelsohn, 1994). Some evidence relates to changes in what teachers believe about their ability to teach and how those beliefs relate to student achievement.

Research instruments and scales have not kept pace with the development and evolution of science education standards, especially regarding teacher self-efficacy beliefs. Previous scales claiming to measure self-efficacy beliefs ranged from designs based on Rotter's (1966) research on locus of control to Bandura's social cognitive theory (Bandura, 1971). Even instruments that focus specifically on preservice science teachers do not include references to the *National Science*

Education Standards (NRC, 1996). If attributes of national science education standards are not explicitly addressed in teacher preparation programs, the gap between expectations and the reality of the classroom will likely widen.

The current study suggests that an instrument measuring self-efficacy beliefs for middle level teachers of science, more closely aligned with the *National Science Education Standards* (NRC, 1996), will provide valuable information to assist the enhancement of existing teacher preparation programs, especially if inservice teachers are considered products of existing teacher education programs. Performance is an important source of efficacy information. Meaningful field experiences in teacher preparation programs promote a positive sense of personal efficacy among prospective teachers, and as teachers gain experience, their sense of personal efficacy increases (Soodak & Podell, 1996).

If teacher self-efficacy beliefs related to standards-based instruction and learning strategies are significant, it can be predicted that beliefs related to student achievement in a standards-based curriculum will be significant as well. An important implication of this change in beliefs about student achievement is an actual increase in student achievement. Teachers who believe their students can achieve at a higher level have classes of students that actually do achieve more (Lawrenz, 1975).

Problem Statements

The first challenge was to determine a set of attributes of standards-based science teaching and learning to be used as the framework for the design of a science teacher efficacy instrument that is standards-based. The next was to develop a valid, reliable instrument, aligned with the *National Science Education Standards* (NRC, 1996), to measure teacher self-efficacy beliefs about teaching science. The third was to determine the relationships, if any, between identified demographic variables, science teacher standards-based self-efficacy beliefs, and teacher beliefs about standards-based student achievement.

Research Questions

1. What are attributes of standards-based science teaching and learning that can be measured with an efficacy instrument?
2. What are inservice teacher standards-based self-efficacy beliefs about their own science teaching?
3. How are inservice teacher beliefs about their own self-efficacy related to their beliefs about standards-based student achievement?
4. How might information resulting from this study be used to inform science teacher preparation programs and strengthen teacher self-efficacy?

Hypothesis

When responding to efficacy items developed from standards-based attributes of science teaching and learning, science teachers will be able to apply critical judgments and decisions about their efficaciousness regarding certain behaviors stated in the items. If teachers have more and varied experiences in both preservice and inservice professional development, then the level of efficacy belief in their own teaching and beliefs about student achievement will both be higher.

Displaying Bandura's Self-Efficacy Design in a Logic Model Schematic

The first step in using a logic model is defining a goal, or **the** goal. For this researcher, the ultimate goal is the improvement of students' science experiences and achievement aligned with national standards. A great deal of what that should look like is identified in national documents describing science education standards (AAAS, 1989; NRC, 1996). Since the design of logic models starts with the goal and works backward, the next steps involve drafting specific, second-level goals for the short term, mid-range, and long term. It is also noted that educational researchers used the same type of model to develop assessment materials and instructional strategies (Wiggins & McTighe, 1998). Within each level of the goals are three components: inputs, activities, and products. Rather than continuing use of terminology favored by business planning, certain education-friendly terms have been adopted. The term "products" become "outcomes"; "activities" is replaced with "actions"; and "inputs" become "resources."

The terms “outcomes,” “actions,” and resources” are common and have wide usage and acceptance throughout the education community. As expected, the term “outcomes” refers to the valued result of improvements in educational practices. “Actions” refers to the identification and elaboration of specific activities that will enable the completion of the desired outcome and includes judgments about the level of knowledge and skills required to complete the activities. The term “resources” refers to any material, person, entity, or other characteristic upon which the people involved can draw to perform the activities and reach the goal. This also includes personal reflection and evaluation about the individual’s capabilities relative to the knowledge and skills needed for the activities.

Bandura's Self-Efficacy Theory in a Logic Model Schematic

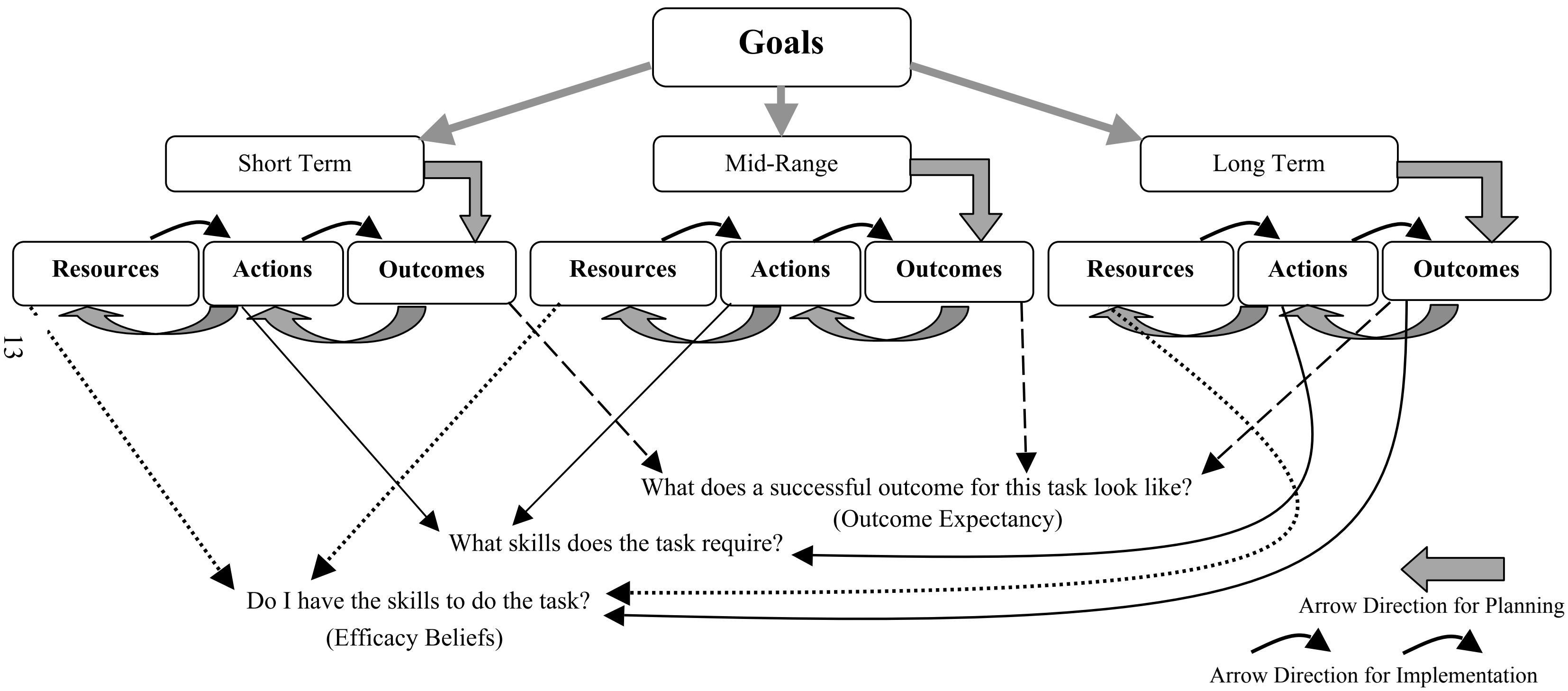


Figure 1.2: Logic model schematic showing connections to Bandura's efficacy beliefs and outcome expectancy

Examples of Resources, Actions, and Outcomes for the Bandura Logic Model Schematic			
Goal Level	**Examples of Resources:**	**Examples of Actions:**	**Examples of Outcomes**
Short-Term	Self-efficacy research literature and instruments previously used to measure teacher efficacy-both general and personal	Identify critical attributes of standards-based teaching and learning Design a measurement instrument for science teacher self-efficacy and beliefs about student achievement whose item are designed in the context of science education standards (K-12) and standards-based science teacher preparation programs	A standards-based science teacher efficacy instrument is reliable and valid
Mid-Range	Self-efficacy research literature and instruments previously used to measure teacher efficacy-both general and personal	Design science teacher education programs that espouse standards-based science Adhere to rigorous, but fair, practices that enhance preservice teacher efficacy and avoid unrealistic idealism	Science teacher self-efficacy is strengthened
Long-Term	Science teacher education programs designed to address standards Inservice professional development sessions that emulate excellent standards-based teacher education programs	Support teachers using standards-based science Provide opportunities for peer support and collaboration Maintain a learning community	Student achievement in science improves

Table 1.1: Examples of resources, actions, and outcomes for the logic model schematic based upon Bandura’s theory.

In his 1997 book, *Self-Efficacy: The Exercise of Control*, Bandura continually reinforces the idea that the strengthening of self-efficacy is not a single-dimension, emotional characterization. He suggested starting with a small, uncomplicated task. If our efficacy is strong, we can approach the task with few misgivings, but if our sense of efficacy about that task is shaky, then we may not risk reinforcing an already tenuous self-belief. If we receive encouragement, for example, then we may try, and we may persevere until we succeed at some level. With our newly enhanced efficaciousness, a second task may not seem so daunting or risky. With each success comes a greater, stronger sense of efficacy and a change in what we think we can do. Bandura states, “They need help to realize that they can be more than they thought they could be” (p. 428).The reference was specifically to people involved in career guidance, but the statement rings true for teachers as well--either about themselves or about their belief in a student’s ability to achieve.

Figure 1.2 shows some of Bandura’s ideas and terminology related to efficacy beliefs and outcome expectancy in the schema of a logic model. It also includes connections to the three questions that are the litmus test of coherence to Bandura’s self-efficacy theories. Table 1.1 provides examples of resources, actions, and outcomes for the logic model schematic based upon Bandura’s self-efficacy theory.

This study adheres to the tenets of Bandura’s explanation (1997) of self-efficacy. The instrument was developed within the domain of science as suggested, and purposefully avoided trying to collect information related to teaching in general, so that teacher responses would be grounded in beliefs and judgments about specific tasks and/or behaviors. As with previous research in the area of teacher self-efficacy in general, and science teacher self-efficacy measurement instruments in particular, the evolution continues as new information and insights become available.

Overview of Instrument Design and Validation

A group of preservice science teachers with questions about being effective was the beginning of the research for this study. A class discussion led to the eventual development of an instrument to measure science teacher efficacy based on standards for science teaching and learning. The informal discussion led to the design of purposeful questions about preservice teacher beliefs about classroom strategies that were aligned with the *National Science Education Standards* (NRC, 1996). The answers to the questions about effectiveness and the preservice teachers' emotional responses during follow-up interviews motivated the desire for a deeper investigation of the relationships between teachers and their beliefs about their ability to teach and their influence on student achievement.

A pilot instrument was designed and administered to preservice science teachers, but the low number of returned, usable surveys prohibited examination by statistical analyses. The usable surveys were then examined by a group of teacher colleagues, gathered for a separate project. They graciously offered a myriad of suggestions about revising the survey. A major change in the subsequent survey was to collect responses from inservice science teachers rather than preservice science teachers. Teacher practitioners have both a greater number and greater variety of experiences in classrooms independent of the intensive support of a mentor teacher or supervisor as was the situation in their internship or student teaching.

Demographic variables were identified. A systematic progression from activities and strategies to tasks to attributes was used to identify specific items for a standards-based science teacher efficacy instrument. Literature and previous research findings were used to guide the identification of the critical attributes and the design of the items.

Both descriptive and inferential statistics were employed in the validation of the instrument--principal component factor analysis, internal consistency reliability estimates resulting in Cronbach's alpha, correlations of significant variables, and multiple linear stepwise regression.

Delimitations, Limitations, and Assumptions

In the design of this study, delimitations related to teaching assignment and grade level were established. One delimitation was to restrict participation to inservice teachers with current assignments as science teachers in which at least 50% of their total assignment was teaching science. The configuration of teaching time was not specified. This parameter permitted teachers with non-traditional daily calendars to participate. A second restriction was that the science teachers must have students in grades 5 through 10.

Several limitations occurred during the study. One of the most significant was that all the data collected was self-reported. Moreover, the teachers volunteered to participate in the study, which may be indicative of a high level of efficaciousness and nonrepresentative of teachers in general. Another limitation was that the data provided by the teachers was incomplete in two ways. An initial set of 250 surveys were distributed, but only 103 were returned, and of those, 87 had complete data sets.

Along with both the delimitations and limitations were two main assumptions made by the researcher. One was the assumption that science teachers were aware of standards in science education that were designed to guide science teaching and learning. The second major assumption was that efficacy is an important factor in teacher quality and student learning.

CHAPTER 2

REVIEW OF LITERATURE

Introduction

The idea of confronting the completion of a task is not new. Our primitive ancestors learned how to size up a situation--we are hungry. Should we chase down this mastodon for dinner? They decided the optimum outcome given the circumstances--we need this food to survive. Then they assessed the likelihood of bringing about the desired outcome--he is big, but we do have these spears, there are several hunters working together, and we need the food to survive. We still size up a situation--my Earth Science class has 32 students; how am I going to survive? We decide the ideal outcome--the students will be able to explain basic processes of how the Earth's surface changes. In addition, we evaluate our capabilities relative to the knowledge, skills, and dispositions necessary to bring about the desired outcome. A major difference, however, is that in modern times, educational psychology has organized the analysis of the process and provided insight into the thinking and learning that is embedded in that process. Two threads of the research, as they pertain to the current study, are what teachers believe about their capabilities for teaching science and how those beliefs are identified and measured.

The literature through which the threads wend is extensive; therefore, this review examines those studies that are direct antecedents of the focus of the current study--the development of an instrument to measure teacher self-efficacy beliefs and their beliefs about student achievement. This chapter begins with an overview of two seminal theoretical positions that have had a significant impact on the development of efficacy instruments: (a) Rotter's (1966) work with locus of control and his ideas about social learning theory, and (b) Bandura's work with social learning theory and his influence on the development of social cognitive theory beginning in the 1970s. The next sections in this chapter examine studies, which focused upon the development of instruments used in measuring teacher self-efficacy, followed by an examination of studies, which focused upon measuring self-efficacy among science teachers.

Self-Efficacy: Defining and Measuring the Construct

Locus of Control

In 1954, Julian Rotter, a clinical psychologist, published *Social Learning and Clinical Psychology,* in which he discussed the integration of learning theory with personality theory. This was a major departure from the prevailing perspective in clinical psychology. Freud's Psychoanalysis held that deep-seated instinctual motives determined behavior; we are not aware of our unconscious impulses, and therapy always necessitates delving into childhood experiences. Drive theory dominated learning approaches. It posited that physiological drives motivated a person to satisfy the drives through various behaviors. These learned behaviors caused a person to either reinforce positive stimulations or avoid unpleasant ones. Rotter posited that behavior is learned through one's interaction with his or her environment, and that personality is a relatively stable set of behaviors for responding to situations in a particular way, but that they can be changed. He believed that since we continue to encounter new situations throughout our lives, our personalities change as a consequence of interactions with a changing environment.

One of the most durable components of Rotter's Social Learning Theory (1966) is the concept of "locus of control." The view of locus of control is that of generalized expectancies for the control of reinforcement. The construct is described as "… a generalized expectancy, operating across a large number of situations, which relates to whether or not the individual possesses or lacks power (or personal determination) over what happens to him" (Battle & Rotter, 1963, p. 482). In Rotter's theory, individuals differ in where they locate the responsibility for what occurs in their lives. The locus of control is along a continuum from totally outside the person (very external), i.e., controlled by fate, luck, or any other powerful force, to totally inside the person (very internal), i.e., controlled by one's own personality, characteristics, or behavior. He was careful to explain that locus of control is not amenable to a typology--an either/or proposition. He also explained that because locus of control is a generalized expectancy, it predicts behavior across situations; however, there may be situations in which a person who tends to be internal may respond more as an external. Battles and Rotter (1963) wrote about teachers who had general beliefs that their level of control over events that happened to them in the classroom were situated along a continuum from completely internal to completely external. That is, believing they had some control over the events (internal), or believing the events were beyond the teacher's individual control and resided outside of the teacher as an individual (external). For example, a science teacher may believe and behave as if student behavior in school is influenced most strongly by what happens beyond the scope of the classroom, but s/he believes and behaves as if s/he has a significant impact on students because they are in her/his class.

Rotter's development of the Internal-External Locus of Control Scale (I-E) attempted to account for and measure individual differences in causal perceptions. Teachers with a higher/stronger sense of internal control generally produce higher levels of achievement among their students than teachers who perceive themselves as having less control, that is, their sense of external control is stronger. The I-E Scale, as developed by Battles and Rotter (1963), was based on a questionnaire used by Bialer in 1961 in the investigation of the development of children's locus of control and conceptualization of success and failure. Bialer's version was in turn based on the adult scale of internal/external control designed by James-Phares in 1957.

The changing personality is another of two main ideas that connect strongly to the current study. Battles and Rotter (1963) did not believe that there was a set time in a person's life when the personality solidified, but he did state that the more experiences you have that reinforce a particular set of beliefs, the more effort and intervention is required to change them. He also believed that people are drawn forward toward their goals rather than being propelled from behind by simply avoiding the unpleasant. This belief included what he called expectancy and reinforcement value.

Expectancy is a subjective prediction of the likelihood that a particular behavior would result in a particular outcome, according to Rotter (1966). If the person has confidence that the outcome is likely, he called it "high expectancy." "Low expectancy" means that the person believes that the outcome is not likely to occur. Since expectancies form from our experiences, we are likely to repeat behaviors that result in reinforcement, and the more reinforced the behavior, the more likely it is to be repeated in future situations.

Rotter (1966) used the term "reinforcement value" to indicate a level of desirability for an outcome of our behavior. For example, if an outcome of a science teacher's behavior is students' returning lab reports, the teacher will repeat past behaviors that resulted in the preferred outcome (more students completing more lab reports) and avoid past behaviors that resulted in the undesirable outcome (students completing fewer lab reports.) Rotter was careful to point out that reinforcement value is highly subjective. For example, in the previous situation of completed lab reports, one teacher may believe that collecting fewer lab reports from students is the desired outcome if completing lab reports interferes with the time spent on interactivity during the lab period.

Reinforcement value and expectancy are intertwined with the idea of locus of control and, since reinforcement value is highly subjective and affected by circumstances, the conjunction is fluid. For example, if the teacher from the previous example uses the lab reports as

20% of the students' grades in each term, s/he has put a value on students' completing and turning in their lab reports--a significant reinforcement value. The teacher may initially have a high expectancy that the lab reports will be completed. If the teacher perceives, however, that because the students are from different backgrounds, the level of influence the teacher exerts varies from one student to another. If the teacher believes the locus of control is internal to him/her and influence over the student is strong, the expectancy may remain high for that student. If the perception is that the locus of control is external to the teacher and other factors exert more influence, the expectancy that the lab report will be completed changes to a lower level. Similarly, if a teacher perceives that factors from outside of school are a stronger influence on students than s/he is, the low reinforcement value and expectancy associated with the completion of lab reports may extend to the whole class or even the entire school, and anything different is reserved for the rare individual student who demonstrates otherwise.

The concepts of locus of control, personality development, and cognitive development continue to exert tremendous influence in studies of teacher self-efficacy. Items in various instruments state options in terms of internal and external control, and interpret findings as levels of influence on teachers' beliefs in their capabilities to teach. The next section begins the discussion of instruments used to measure these levels of perception of level of control.

Social Cognitive Theory and Efficacy

In 1971, Albert Bandura, a cognitive psychologist, published a social cognitive theory, as a component of social learning theory that ran counter to the doctrinal sense of determinism. He used the word *determinism* to "… signify the production of effects by events rather than in the doctrinal sense…that actions are completely determined by a prior sequence of causes independent of the individual" (Bandura, 1997, p. 7). According to determinism, there is no choice of actions based on information derived from, and analyzed after, a particular event. The influence of cognitive processes on subsequent actions or behavior is absent. Another strong aspect of the deterministic school of thought is the operant view set forth by B.F. Skinner. His perspective was that people are repositories for past stimulus inputs and conduits for external stimulation. Bandura supported the idea that "…people are at least partial architects of their own destinies. It is not the principle of determinism that is in dispute, but whether determinism should be treated as a one-sided or a two-way process" (p. 8). If one subscribes to this perspective, the implications for teaching and learning are tremendous. In the operant view, an individual is capable of responding to stimuli from the environment but cannot change the environment--the causation is unidirectional. For example, according to the operant view of determinism, if a science teacher has a student who needs adaptive instrumentation to participate in a physics lab, the teacher could not be creative in figuring out a way to enable the student's participation. The only possibilities for achieving the goal of the student's participating would be if the teacher had previously encountered a similar student or had some other experience from which s/he could extract the solution.

Bandura (1971), however, presented a case for bidirectionality. He believed that people attain the results they desire by developing cognitive guides, by attaining self-described goals, by making choices about existing environments, and by creating or changing environments that do not suit their purpose. The degree of success a person attains is affected by the means of self-influence, amount and quality of foresight, and level of proficiency the person has. Higher levels of these qualities result in greater success in achieving what a person wants. This concept is crucial in the study of self-efficacy beliefs.

Bandura (1971) defined perceived self-efficacy as "... beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (p. 83). He studied efficacy among various groups of individuals, including classroom teachers, athletes, addicts in treatment, and aggressive adolescents. This broad application and interpretation of his theories led Bandura along a scholarly path of continually developing, describing, and refining his ideas. In response to other researchers' efforts to add to the construct description, Bandura (2001) also addressed misconceptions he saw emerging from others' research, and he created a guide to developing efficacy scales. Since the measurement of teacher self-efficacy is the focus of this dissertation research, that thread will be followed through Bandura's work.

The first aspect of self-efficacy beliefs, the degree to which one believes in one's capabilities, is a factor of level, generality, and strength. The second aspect of self-efficacy is related to the level of skills necessary to complete the task and involves the capability of self-evaluation. The third aspect is the expectation of a particular outcome when one has the belief in oneself and knowledge of a degree of one's skillfulness. This expectation of a particular outcome was labeled *outcome expectancy*. Where efficacy beliefs are connected to a person's conviction that s/he has the necessary knowledge and skills to perform a specific task, outcome expectancy relates to the person's expectation of certain consequences if the task is performed competently (Bandura, 1986). These two factors--self-efficacy beliefs and outcome expectancies--are the foundation of the construct, and they have resurfaced throughout efficacy research.

In 1997, Bandura published *Self-Efficacy: The Exercise of Control,* a thorough compilation of his theories, findings, and conclusions from more than 30 years' research. In it, he built a strong case for the importance of a teacher's sense of efficacy. "Teachers' beliefs in their instructional efficacy is a much stronger predictor of the academic attainments of younger students than of older students" (p. 242). He also emphasized a particular concern for efficacy among science teachers, given the importance of scientific literacy and competency during the current technological transformations. In the continuing discussion, Bandura made two especially important points. The first is that "… teachers' sense of instructional efficacy was the best predictor of commitment to the teaching profession" (p. 243) above a school climate of collegiality and support, salary, and teaching experience. The second is that "Teachers' instructional efforts are governed more by what they believe they can accomplish than by their view of other teachers' abilities to prevail over environmental obstacles by effective teaching" (p. 243). For example, a science teacher believes s/he can overcome the shortage of adequate middle level Earth Science textbooks by supplementing lessons with activities and projects while, at the same time, s/he believes most other teachers might not be so willing to work so hard to improve their students' achievement in Earth Science.

Bandura (1997) also was concerned about the measurement of efficacy, and he put forward three challenges to effectively measuring it. The first challenge is determining different levels of efficacy related to the same task. A teacher may believe s/he knows Life Science concepts about interdependence adequately enough to teach fourth- or fifth-grade students, but doesn't feel s/he understands them deeply enough to teach the concepts to tenth graders. A second challenge is that efficacy beliefs differ in generality. That is, the efficacy beliefs might be generalized across domains or may be limited to specific circumstances. For example, a teacher may believe s/he can answer questions about big ideas in physics, because s/he is confident in her conceptual understanding of the phenomena. On the other hand, s/he may be unable to apply the correct formula to a given problem, because s/he doesn't believe s/he can do the mathematics. The third challenge is determining the strength of the efficacy beliefs. Teachers whose efficacy beliefs are weak are likely to give up more quickly when they encounter challenges. Those whose beliefs are more tenacious will persevere through considerable obstacles. Thomas Edison's belief that he could create a workable incandescent light bulb was durable enough to withstand failures in dozens of experiments conducted in the search for the best filament.

In addition to expressing his views on measuring efficacy, Bandura (2001) also developed guidelines for creating efficacy scales. He stressed that a major component of item design is using language that indicates *can do* rather than *will do*. Since efficacy is a belief of what a person is *capable* of doing, not a belief of what a person is *willing* to do, it is important to make this distinction. This point became very important as efficacy research moved into education in a major way. Some researchers continued to follow Rotter's thinking about where teachers believed the capability emanated from. They developed measures based on internal versus external control, while others followed Bandura's lead, using the two factors of self-efficacy beliefs and outcome expectancies in the development of other measurement instruments. The following section begins the examination of various scales developed to measure efficacy in both teacher education, in general, and science teacher education, in particular.

Measuring Efficacy in Teacher Education

The Rand Study

In mid-1975, the Board of Education of the Los Angeles Unified School District decided to study gains in sixth grade reading achievement in selected schools with predominantly minority populations. Three years prior, the district installed the School Preferred Reading Program, and they wanted to gauge its success. In January 1976, the Board contracted with the Rand Corporation of Santa Monica to analyze the program's effects on reading progress among Black and Mexican-American students. In August of 1976, the researchers, Armor et al., published a report that would have widespread implications throughout efficacy research.

The Board of Education wanted information about school leadership, classroom atmosphere, teacher attributes, and the content and implementation of the reading program. They also specified that the expression of reading achievement outcomes was to be in terms of scores on standardized tests. Doubts were expressed among the research team about the accuracy of such scores as indicators of true gain, but they also acknowledged that the use and acceptance of the scores were widespread. Thus, the researchers set out to discover which classroom approaches and which school policies supported the recorded gains. In order to ascertain which policies and practices led to gains in reading scores, the team had to determine if the gains were valid and reliable, identify characteristics of the schools and programs that could account for the gains, and provide recommendations for some possible ways to extend the successful policies and practices to other schools across the district. They looked for the factors in the reading program and environment that could be replicated most easily.

The researchers selected a sample of 20 elementary schools that had all shown significant and consistent reading gains for the sixth grade on national standardized tests between 1972 and 1975. The student population fit the criterion of Black to Mexican-American ratio (about the same number of each in every school in the study), the schools' location in low-income neighborhoods, and an enrollment of at least 400 students in each school. Because of the possibility of wide differences between the sixth-grade students' reading scores in two successive years due to differences in socio-economic status, year-to-year data can be misleading. To counter this problem, the research design team decided that a longitudinal study of students would be more appropriate in this case, rather than using data from the sixth grades in each of the years under scrutiny. Since all the records of academic achievement, as well as other demographic information, were stored in a central location, the researchers were able to retrieve the data on reading scores for the students enrolled in the sample schools. They had access to records from the students' third and fifth grade reading tests as well as data from the junior high schools in which the students had enrolled after spending sixth grade in one of the sample schools. They were able to trace the records of at least 68% of the students in every one of the sample schools.

Information in the categories other than reading scores were collected through personal interviews or self-administered questionnaires, and in a 3-school subset of the 20-school sample, intensive case studies were conducted. The questionnaire was administered to 81 of the 83 teachers who had taught sixth grade in the designated schools in 1974 and 1975, and all principals and reading specialists present in 1976 were interviewed individually. With this context established, the following paragraphs focus on the aspect of the Rand study that is connected to the current measurement of efficacy among teachers.

One relatively small portion of the comprehensive Rand study (Armor et al., 1976) measured a single aspect of teachers' attitudes: their sense of efficacy in dealing with minority students. Two items on the teachers' questionnaire were based on Rotter's work from 1966. The first question asked if the teacher felt "that 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" (p. 23). The second asked whether the teacher thought that "if I try really hard, I can get through to even the most difficult or unmotivated students" (p. 23). The teachers' responses were coded and transformed into quantitative data for analysis and comparisons. The term *general teaching efficacy,* or GTE, was assigned to the score of the response to the first question, and *personal teaching efficacy* or PTE, was the score of the response to the second question. The responses were combined into a single measure of efficacy indicating the extent to which the teacher believed s/he could produce an effect on the students' learning and was called *teacher efficacy* (TE).

One particular aspect of the study's findings supports a strong relationship between the teachers' sense of efficacy and the positive achievement of their students. Specifically, the more efficacious the teachers felt, the more their students advanced in achievement, and that "… the teacher *attitudes* are more significant than their background characteristics" (p. 37). Since the study of teachers' efficacy was not the central purpose of the research, it was not possible nor feasible to determine the best methods to raise teachers' sense of efficacy. What did emerge was the attention to the two questions as somewhat of a baseline instrument representing Rotter's concept of locus-of-control. The next section examines a second major interpretation of the construct of self-efficacy and the connections to teachers and teaching.

Teacher Locus of Control (TLC) Scale

In 1981, Rose and Medway undertook a task to advance the application of Rotter's ideas of locus of control among classroom teachers. They point out that the I-E scale, developed by

Rotter (1966), was used almost exclusively by researchers examining the relationship between teacher perceptions of control and student achievement. Rose and Medway also state that when the I-E scale was used to measure a level of specificity consistent with classroom-level expectancies associated with teaching, the role was contrary to the intent and design of the I-E scale. It also was not intended to be highly predictive of classroom-level process variables and specific teaching outcomes, such as student achievement. In following the recommendation of other researchers to develop scales based on specific situations and tailored for specific populations, Rose and Medway developed the Teacher Locus of Control (TLC) Scale to measure elementary school teacher perceptions of control in the classroom. They planned to obtain separate scores for success and failure situations because teachers attribute causality depending on the nature of the classroom and performance outcomes of the students.

The TLC was based on a questionnaire-- the Intellectual Achievement Responsibility (IAR), developed by Crandall, Katkovsky, and Crandall. Like the IAR, the TLC was a forced-choice instrument with items that are roughly equivalent in strength. Unlike the IAR with 62 items, the TLC consisted of only 28 items. The forced-choice items required teachers to mark an option that indicated either external or internal control of various classroom events, and one point was awarded for each internal alternative. The items relating to internal control were also tagged to provide additional information in the analysis. Items marked with I+ indicated belief of internal responsibility for student success; I- items indicated belief of internal responsibility for student failure. The initial set of items was constructed with reviews and editing help, depending on how the item fared in successive administrations of the instrument. Controversial items were either modified or removed. After the administration of each of the versions of the scale, Rose and Medway (1981) made decisions to retain items if they clustered together on their respective subscale, and if they produced significant biserial correlations ($p<.05$) of the item with the total subscale score with that item deleted. One of the findings from the study was that Rose and

Medway concluded that teachers with higher scores seemed more likely to accept personal responsibility for classroom events.

Teacher Locus of Control Scale Validation

The TLC scale represents the final step in a series of five studies, each of which involved different groups during the development and validation of the instrument. Four other groups were administered three earlier versions of the scale. The total number of teachers in the first three groups was 183, and the final group consisted of 89 female fourth-grade teachers in a large school district (> 50,000 students) for a total of 272 participants. Responses of the 89 teachers to the final set of items, 14 items describing positive or success situations and 14 describing negative or failure situations, were analyzed using a principal-components analysis with an iterative procedure that replaced the diagonal elements of the correlation matrix with improved estimates of communalities. The factors, limited to two, were rotated with a varimax procedure to simplify interpretation of the factor pattern. The analysis clearly yielded two subscales: I- (Factor 1) and I+ (Factor 2), with appropriate items on each. In reporting on this and all other administrations of the TLC in this study, no information specific to teachers' actual responses or range of scores is included. The authors focused on the data analyses' results, interpretations, and conclusions of the studies.

Two groups of teachers completed Rotter's 1966 I-E scale and the Crowne-Marlowe Social Desirability scale in addition to the TLC scale. The scores for the group of 89, fourth-grade teachers showed a Pearson-product moment correlation of -.22 ($p < .05$) for the I-E and I+ scores, and the correlations for I-E and I+ scores were -.32 ($p < .05$) for the second group of 45 elementary teachers. Rose and Medway (1981) reported non-significant correlations were found between I-E scores and I- scores for the same groups of 89 teachers and 45 teachers. From the data of the group of 89 elementary teachers, Rose and Medway also reported that all corrected-item correlations were significant ($p < .01$) and that the Kuder-Richardson formula 20 reliabilities were .81 and .71, respectively, for the I- and I+ subscales.

The mean scores for the 89 teachers completing the TLC were 6.89 ($SD = 3.66$) for the I- subscale and 8.07 ($SD = 2.98$) for the I+ subscale. Differences between groups of teachers were examined using *t*-test comparisons--no significant differences were found between black and white teachers and between teachers in urban and suburban schools. Teachers in low SES schools, however, scored significantly higher on the I+ subscale but not the I- subscale compared to teachers in high SES schools. The mean scores for low SES and high SES schools on the I+ subscale were 9.15 and 7.41, respectively ($p < .01$, $t = 2.78$), but the scores for the low SES schools compared to the high SES schools were closer together ($m = 7.58$ and $m = 6.46$, respectively) with $p < .01$ and $t = 1.38$. This seems to indicate that, in schools serving a population of students from families with lower socio-economic resources, teachers may see themselves as having a greater influence on students who succeed than on those who do not succeed. It may also indicate that teachers in such schools recognize that the influence of community situations and events on students who are not successful in schools is greater than the influence of teachers.

Rose and Medway (1981) refer to the study conducted by Berman, McLaughlin, Bass, Pauly, and Zelman that examined factors affecting the implementation and continuation of federally funded educational projects. “They found evidence that teachers who believe they have the capacity to affect student performance tend to continue to apply project methods and materials after federal support has terminated” (p. 186). Rose and Medway suggest that the prediction of project implementation behavior could be used as a criterion of TLC score validity. The TLC Scale was administered to 45 elementary teachers who had recently been involved with substantial professional development aimed at encouraging them to adopt new instructional strategies. Several behaviors were identified as indicative of the strategies presented in the professional development sessions. Rating choices were expressed according to how often the teacher engaged in the activity. The range of choices included the following: (a) daily, (b) once a week, (c) always, (d) sometimes, or (e) never. The report did not address the issue of overlap that seems apparent in the response choices. An analysis of the results showed that the TLC Scale was related to implementation of several program components, most often in the middle elementary grades, whereas the I-E scale was not.

Rose and Medway (1981) conclude that the TLC Scale was more predictive of teacher and student behavior than Rotter's I-E scale. The TLC Scale was internally consistent and yielded higher correlations with classroom teaching behaviors. The classroom behaviors they found "… to be more characteristic of internal teachers (i.e., fewer disciplinary commands given to students, lower rates of inappropriate student behavior, and higher rates of student self-directed activity) are also those that maximize instructional efficiency" (p. 188), and internal teachers tend to employ improved educational practices more often than external teachers do. They attribute this difference partly to the fact that the TLC items are written in language that is more specific to classroom situations, where the I-E scale is a measure of more generalized beliefs.

A second validity study (Rose & Medway, 1981) was an examination of the relationship between TLC scores and classroom behavior for a representative sample of 30 female fourth-grade teachers. Observers recorded teachers' use of disciplinary actions, techniques for holding students accountable for their performance, and teacher engagement in instructional activity. The student behaviors observed were self-directed activity, attending or listening, and appropriate behavior.

Findings (Rose & Medway, 1981) for the I+ subscale indicated that highly internal teachers in low SES schools gave fewer disciplinary commands to students. In high SES schools, high I+ teachers held fewer students accountable by selecting non-volunteers to answer questions and had students who engaged more frequently in self-directed activity and less frequently in attending and listening behaviors. Teachers in high SES schools scoring as internals on the I-subscale had fewer students engaged in inappropriate behavior. For the I-E scale, only one significant correlation appeared -- high SES students in classrooms of external teachers engaged more often in attending behavior.

Recommendations From TLC Scale Studies

Several recommendations emerged from the studies using the TLC Scale. The first was that professional development activities should enhance teacher control beliefs and a sense of internality prior to introducing innovative instructional methods. A second recommendation was that additional studies should be conducted to determine the degree of relationship between teacher locus of control and an indicator of effective teaching, such as student achievement. They also recommended that additional studies be directed toward identifying the antecedents of teacher locus of control, such as confidence in teaching ability, attitudes expressed by colleagues and administrators, the relative stability of student populations, and prior exposure to predictable and consistent standards of teacher behavior.

Teacher Efficacy Scale (TES)

In 1984, Gibson and Dembo published the results of a study they designed and conducted to develop an instrument that would measure teacher efficacy. They wanted to provide construct validation for the variable and examine the relationship between teacher efficacy and observable teacher behaviors. The study stretched over three phases of instrument development and involved over 200 teachers, both preservice and practicing. The first phase, factor analysis, involved the development of a pool of items that could potentially be included in the Teacher Efficacy Scale, refinement of the items, and the administration of the TES to ascertain the emergence of identifiable factors.

Phase 2 was a multitrait-multimethod analysis. The traits compared across the methods were efficacy, verbal ability, and flexibility; each trait was compared across each method and to other traits in the analysis. The methods used were: (a) closed-ended and (b) more open-ended. The closed-ended items of the second phase belonged to the Teacher Efficacy Scale from Phase 1. The list of 30 items offered choices from 1-6 on a Likert-type scale. The items that were more open-ended consisted of a list from which teachers were asked to check 10 of 20 variables they believed contributed the most to student success or failure in school. Half the items were teacher-related variables, such as teacher ability to individualize instruction and teacher management skills, and half were related to external variables, such as parental support and student home environment.

The purpose of the third phase (Gibson & Dembo, 1984) was to gather information about the possibilities inherent in a methodology that was presented as if it were a stand-alone study, and the final discussion drew information from each phase and included suggestions for further study. A more detailed review of each phase appears in subsequent paragraphs.

Validation of the Teacher Efficacy Scale

Phase 1: Factor analysis. Before the first phase of the study could begin, Gibson and Dembo (1984) conducted a pilot study to generate items. These sample items were based on interviews with teachers about what they thought were characteristics of effective teachers. They also were derived from an analysis of the literature that reported characteristics of teachers that previous researchers identified as having a sense of self-efficacy. For example, persistence in helping struggling students tended to be more evident in teachers with high self-efficacy and responding to students with positive feedback rather than criticism when the student gives a response was also more evident. The teachers in the pilot study had the opportunity to respond to the pool of 53 pilot items, after which Gibson and Dembo used a principal components factor analysis and eliminated items with poor variability, and maintained items that loaded clearly on one of the substantive factors. Any items remaining in the pilot instrument were clarified and checked for proper item construction. The initial 53 items resulted in a set of 30 items in a Likert-scale format. The choice for each item was to select from numbers 1 to 6 (strongly disagree to strongly agree) to indicate the level of agreement with the item statement.

With the first instrument ready to be administered, 208 elementary teachers were selected from thirteen K-6 schools within two neighboring, unified school districts. Their years of teaching experience ranged from 1 to 39 years. The report of Gibson and Dembo's (1984) study stated that the teachers' years of teaching experience were divided into the following increments: 26% had 1-5 years, 25% had 6-10 years, 23.7% had 11-20 years, and 16.3% had 21-39 years. The percentages sum to 91% and no accounting is given for the remaining 9%. Of the 208 teachers, 75% were female. The teachers completed the 30-item instrument, the Teacher Efficacy Scale (TES), after it was presented at regularly scheduled faculty meetings, and most were completed within 15 minutes.

The data analysis began by using SPSS for the factor analysis, used squared multiple correlations in the main diagonal of the correlation matrix as communality estimates, and used iteration procedure for improving estimates. From this preliminary solution, two factors were extracted that were concurrent with Bandura's two-factor model of self-efficacy. Later, both oblique and orthogonal rotations were used to compare item loadings and degree of correlations between factors. With the delta value set at zero, an oblique rotation revealed that the two factors were only moderately correlated ($r = -.19$). The same items showed significant loadings for both solutions, and the orthogonal rotation was used as the final solution. The decision to include an individual item rested with its showing factor loadings greater than or equal to .45 (Gibson & Dembo, 1984).

With this criterion established, two factors emerged from the data. Factor 1 (9 items) accounted for 18.2% of the total variance, and Factor 2 (7 items) accounted for 10.6%. Each remaining factor accounted for less than 6% of the total variance. Factor 1 appeared to represent a teacher's sense of personal teaching efficacy (the belief that the individual had the skills and abilities to bring about student learning) and was labeled Personal Teaching Efficacy (PTE). All items included in Factor 1 reflected the teacher's sense of personal responsibility in student learning and/or behavior and corresponded to Bandura's self-efficacy dimension. This dimension was also reflected in one of the items used in the Rand report (Armor et al., 1976): "If I try really hard, I can get through to even the most difficult or unmotivated students" (p. 23). Personal efficacy also bears a resemblance to a component of Rotter's (1954, 1966) work called internal control (I+), which is discussed in an earlier section of this report. The following is an example of items that loaded for Factor 1 on the TES:

> Self-efficacy (Personal Teaching Efficacy)
> When a student is having difficulty with an assignment, I am usually able to adjust it to her/his level.
>
> If one of my students couldn't do a class assignment, I would be able to assess accurately whether the assignment was at the correct level of difficulty (p. 573).

Factor 2 of the TES, General Teaching Efficacy, reflected the teacher's belief about the general relationship between teaching and learning. It was represented by the second item in the Rand (Armor et al., 1976) study: "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." (p. 23). Gibson and Dembo (1984) state, "This second factor clearly corresponds to Bandura's outcome expectancy dimension" (p. 574). As with the previous example, a resemblance is indicated between Gibson and Dembo's work about to TE to Rotter's (1954, 1966) external control factor (I-). Examples of items loading on Factor 2 Outcome Expectancy (General Teaching Efficacy) of the TES are the following: (a) The amount that a student can learn is primarily related to family background and (b) If students aren't disciplined at home, they aren't likely to accept any discipline (Gibson & Dembo, p. 573).

An analysis of internal consistency reliability yielded Cronbach's alpha coefficients of .78 for the personal teaching efficacy factor, .75 for the teaching efficacy factor, and .79 for the total 16 items. Because acceptable reliability coefficients resulted from only 16 of the original 30 items, the authors suggested further research using 16-20 items, rather than using just the 16 items that showed significant loadings on the two factors.

The authors determined that the results of Phase 1 clearly supported Bandura's (1977) two dimensions of self-efficacy--Factor 1 represents self-efficacy (Personal Teaching Efficacy) and Factor 2 represents outcome expectancy (Teaching Efficacy). In support of the connections made between their developing instrument and Bandura's definitions, Gibson and Dembo (1984) state, "…outcome expectancy would essentially reflect the degree to which students can be taught given their family background, socio-economic status (SES), and school conditions" (p. 574). With teaching efficacy and outcome expectancy, the researchers, however, do not make a connection to the aspect that Bandura emphasized, namely the judgment of relationship between the individual's behavior and the outcome**.** "An outcome expectation is a judgment of the likely consequences such performance will produce" (Bandura, 1997, p. 21). Gibson and Dembo's emphasis on a student's ability to learn does not seem to equate with Bandura's conceptualization of an individual's judgment regarding outcomes. The researchers' connections between their results related to personal teaching efficacy and Bandura's dimension of self-efficacy are more closely aligned, because the items are designed to elicit a judgmental response about a teacher's own skills and abilities. Gibson and Dembo concluded that teacher efficacy, as measured by the Teaching Efficacy Scale, is multidimensional, and is comprised of at least two factors that are clearly distinguishable from each other. They also recommended further investigations into teacher efficacy because the two subscales showed only a moderate correlation.

Phase 2: Multitrait-multimethod analysis. During the course of Phase 1, Gibson and Dembo (1984) established that their new instrument held the very real potential for measuring teacher efficacy. During Phase 2, their intention was to create a 2 x 3 matrix and identify any statistically significant relationships between the two methods (closed-ended and open-ended) across three traits (teacher efficacy, verbal ability, and flexibility) or between any two traits. The modified Teacher Efficacy Scale (TES) was administered again, this time to 55 veteran teachers enrolled in a graduate education course. Each participant responded to items on instruments related to three constructs: teacher efficacy, verbal ability, and flexibility. Gibson and Dembo documented detailed information about the origins of the instruments for verbal ability and flexibility, but provided no insight about why these two traits were selected over any others. Only a vague reference that the traits were previously identified as characteristics of effective teachers was discerned. Table 2.1 displays the traits and the instruments that were part of Phase 1 and Phase 2.

Comparison of Instruments

Test Name	Trait	Test Description
Instruments used in Phase 1		
1. Teacher Efficacy Scale 2. Unnamed open-ended measure	Efficacy	1. Closed-ended-30 items with choices from 1-6 on a Likert-type scale. 2. Somewhat open-ended-teachers to choose 10 of 20 items that contributed most to student success or failure in school.
Instruments used in Phase 2: Beginning Teacher Evaluation Study, Phase 2, 1973-76: Teacher Aptitude battery		
1. Verbal Facility Test 2. Controlled Associations Test	Verbal Ability	Somewhat open-ended 1. Thirty items-participants asked to choose the best of five words to complete a given sentence. 2. Eight items-each item asked the respondent to write as many synonyms as possible for each of the stimulus words.
1. Finding Useful Parts 2. Planning Test	Flexibility	More open-ended 1. Ten items-each item had five options from which the respondent could choose the object that could be used as a makeshift substitute for a missing original object. 2. A test of sensitivity to problems-practical problems are presented to the respondent, as well as several plans for solving the problem; the respondent is asked to indicate what is wrong with each of the proposed plans.

Table 2.1: Traits and instruments used in Phase 1 and Phase 2 of Gibson and Dembo's (1984) study.

One component in Phase 2 different from Phase 1 was the inclusion of items that were more open-ended. Teachers were asked to check 10 of 20 variables they thought contributed most to a student's success or failure in school. Ten items were teacher-related variables, such as the level of teacher rapport with students, a teacher's ability to individualize instruction, and teacher management skills. The other 10 items were externally related variables, such as student intelligence, parental support, and student home environment (Gibson & Dembo, 1984).

In Phase 2 (Gibson & Dembo, 1984), data were collected from the teachers in four consecutive class sessions in this order: (a) the Teacher Efficacy Scale, (b) the open-ended efficacy measure and the Verbal Facility Test, (c) the Controlled Associations Test and the Finding Useful Parts Test, and (d) the Planning Test. All were administered according to acceptable professional standards. In order to assure all possible combinations of traits and methods, the authors created a 2 x 3 matrix. They used the three traits of teacher efficacy, verbal ability, and flexibility, across two methods (closed-ended and open-ended.) A multitrait-multimethod analysis was used, and correlations of variables within and between methods were computed to examine convergence of the construct across methods and to determine discriminability of teacher efficacy from other constructs.

The results (Gibson & Dembo, 1984) indicated that both methods producing data on teacher efficacy converged, as indicated by a positive correlation of .42 ($p < .001$) for the two methods. All three traits pass the criteria for convergent validity by showing values significantly beyond the .05 level (.30, .39. and .42 for verbal ability, flexibility, and teacher efficacy, respectively). Establishing discriminant validity, which was the next task, occurred in two steps. First, each validity value was compared with all values in the heterotrait-heteromethod cross-block to determine whether the correlations between different methods of measuring the same trait exceeded correlations between the trait and other traits not having that method in common. In all cases, the correlation coefficients of each trait measured against both methods exceeded the correlations of one trait to the other two. The second step in determining discriminant validity was to compare each trait's validity value with values in the heterotrait-monomethod cross-block in which that trait was involved. The purpose of this step was to determine whether the correlation between different methods of measuring the same trait exceeded correlations between that trait and other traits that had the method in common.

The results (Gibson & Dembo, 1984) were mixed but supportive-there was a shared method variance when three traits were measured by the closed-ended method compared to the open-ended method, and the trait variance for teacher efficacy either exceeded the method variance or was independent of the method. The same pattern of trait interrelationship showed in all of the heterotrait cross-blocks of both heteromethod and monomethod blocks, but it was not definitive because of small differences. There was considerable support for the convergence of teacher efficacy when measured by two different approaches and discriminability from other constructs already in use. The results verified the distinction between efficacy and two other constructs (verbal ability and flexibility) already identified in research as characteristics of effective teachers, which may be the rationale used by Gibson and Dembo for the inclusion of these two constructs in their study.

Phase 3: Classroom observation. The third portion of Gibson and Dembo's (1984) study was designed as more of a pilot study to gather information about the generative and analytic possibilities of a more qualitative methodology. To determine which teachers would participate in Phase 3 of the study, Gibson and Dembo examined the scores of all the respondents for both Factor 1 and Factor 2, and factor scores were computed for all the teachers who participated in Phase 1. High scores on Factor 1, Personal Teaching Efficacy, indicated high personal efficacy, which was comparable to Bandura's dimension of self-efficacy. A low score on Factor 2, General Teaching Efficacy, indicated high teaching efficacy, which was comparable to Bandura's dimension of outcome expectancy. By examining the scores at both extremes of each subscale, researchers generated a pool of possible participants by selecting teachers whose combined scores indicated high self-efficacy (PE) and high outcome expectancy (TE) or low self-efficacy (PE) and low outcome expectancy (TE), and then assigned a single indicator for each teacher. Sixteen teachers selected from the sample had scores at the extreme ends of each scale-the top 6% of the frequency distribution for Factor 1 and the bottom 22% of scores for Factor 2. Researchers requested permission to observe the classrooms of all 16 teachers and, of the 12 who granted permission, eight were selected if they had the most discrepant and equally distributed scores. All eight teachers were from two schools which minimized the influence of organizational variables.

Seven observers were trained to collect data from video tapes made during 7.5 hours of observations of instructional time collected from each participating teacher. The interrater reliability for the seven observers ranged from .73 to .91.

One of the qualitative observation strategies was teacher-use-of-time (Gibson & Dembo, 1984). The observer coded a teacher's behavior whenever activities were introduced or changed by the teacher, and it measured a ratio of time spent on teaching and academic learning compared to non-academic activities. The data from the teacher-use-of-time instrument was analyzed using the total minutes allocated by the teachers to each of the observation categories: (a) daily rituals, (b) transitions, (c) whole class, (d) small group academic, (e) checking seatwork, (f) preparation or paperwork, (g) intellectual games, (h) unfocused small talk, and (i) recess. Total academic time was calculated by adding together the total number of minutes for whole-class activities, small group academic work, and checking seatwork for each group. A total for the non-academic time was the sum of the remaining categories: daily rituals, transitions, preparation or paperwork, intellectual games, unfocused small talk, and recess.

Another strategy used by the observers involved a question-answer-feedback sequence (Gibson & Dembo, 1984). The observer then coded the teacher and student dyadic behavior during question-and-answer interchanges, and the quality of the student's response and the nature of the teacher's feedback were coded. For example, praise-per-successful interaction was derived by dividing the number of teacher praises by the number of correct student responses. A corresponding derivation represented the number of teacher criticisms divided by the number of incorrect student responses. A similar process occurred for teacher persistence following a student's incorrect response. The observers recorded teacher behaviors such as repeating the question, providing a clue, or allowing another student to call out the answer when a student answered incorrectly or not at all.

Recording and analyzing the question-answer-feedback sequence was more complex (Gibson & Dembo, 1984). It involved examining the question the teacher was asking, the type of initial response provided by the student, and the type of feedback given by the teacher after the student's initial response. The measures of teacher praise were controlled for absolute frequency differences in students' correct and incorrect responses, and the measure of teacher praise was derived by dividing the number of teacher praises by the number of correct student responses. This praise-per-successful interaction indicates frequency of rewarding behavior. The number of teacher criticisms was divided by the number of incorrect student responses given, and teacher persistence was determined as a ratio of feedback interactions in which a teacher repeated the question, provided a clue, or asked a new question following a student's incorrect response. The teacher's lack of persistence in failure situations was defined as a ratio of feedback interactions in which a teacher gave the answer, asked another student, or allowed another student to call out the answer when the initial student failed to respond or responded incorrectly. Percentage figures were derived from the question-answer-feedback sequence instrument in which absolute frequency differences across teachers was statistically controlled to allow comparison of relative differences between groups. One-tailed *t*-tests with the teacher as the unit of analysis were used to analyze differences between high-and low-efficacy teachers in behaviors related to academic focus, teacher feedback, and teacher persistence.

Upon examination of the teacher-use-of-time variables when combined into academic and non-academic totals, there was no significant difference between high- and low-efficacy teachers; however, patterns emerged upon inspection of the academic subcategories (Gibson & Dembo, 1984). For example, a significant difference was noted in the amount of time spent in small-group instruction $t(6) = 2.23$, $p < .05$; low-efficacy teachers spent nearly half of their observed time in small-group instruction, while the high-efficacy teachers spent approximately one-fourth of their observed time in small groups. Although not statistically significant, it was noted that high-efficacy teachers spent approximately 50% more time in whole group instruction and more time monitoring and checking seatwork than low-efficacy teachers.

Gibson and Dembo (1984) reported relationships between several sets of variables. Significant details were provided about the analysis and results. Within non-academic categories, high- and low-efficacy teachers spent roughly equivalent amounts of time doing similar activities, except for using intellectual games and preparation of paperwork. Low-efficacy teachers spent only about 2% of their time doing intellectual games, but no games of similar types were observed in the classrooms of the high-efficacy teachers. In addition, high-efficacy teachers tended to spend more time than their counterparts did in the preparation of paperwork.

As with the recording of the question-answer-feedback sequence, the data analysis and results are more complicated. Praise and correct student responses were recorded and then displayed as ratios. The teacher feedback for incorrect student responses was in the form of criticism, persistence, and/or lack of persistence. They also were recorded and then reported as ratios. Gibson and Dembo (1984) reported a significant difference between high- and low-efficacy teacher criticism to incorrect student responses, $t\ (6) = 5.17$, $p<.01$. They found that 4% of low-efficacy teachers' feedback to incorrect answers was in the form of criticism. Although this figure may not seem significant, no observations of criticism occurred among any of the high-efficacy teachers under the same circumstances.

In the categories of praise-per-correct response and persistence-per-incorrect response, Gibson and Dembo (1984) report only a slight difference between high- and low-efficacy teachers, with the mean difference being slightly in favor of the high-efficacy teachers. In the final category of lack of persistence-per-incorrect response, the low-efficacy teachers were more likely to continue through the lesson by giving the answer, asking another student, or allowing another student to call out before the first student gave the correct answer, $t(6) = 3.29$, $p<.01$. Both high-efficacy and low-efficacy teachers provided other opportunities for students to correct their responses (persistence), but high-efficacy teachers were more effective in leading students to correct responses through questioning. Low-efficacy teachers would go on to other students or another question (lack of persistence) before the initial student could arrive at a correct response.

Academic focus. At least one previous study asserted that successful teachers spend less time on nonacademic activities and maintain a strong focus on academics (Rosenshine, 1978). Gibson and Dembo (1984) attempted to examine global academic time by collapsing the categories of the teacher-use-of-time measure but produced no significant results. They believed that academic time coding of the activities observed during the study might be more of a reflection of allocated time and organization rather than of academically engaged time or academic focus. For example, they posited that time allocation for planned instruction, enacted instruction, and student time spent engaged in learning may all be very different and may produce different data. Another significant idea they considered was that coding in the teacher-use-of-time measure did not take the level of student engagement in the instructional activities into account. Anecdotal observation data suggested that the level of student engagement is different between high-efficacy and low-efficacy teachers.

Student grouping. Several findings about student grouping emerged from Phase 3 of the study (Gibson & Dembo, 1984). High-efficacy teachers allocated approximately one-fourth of the total observed time to small-group instruction--an average of 124.8 minutes of the total observed time of 134 minutes (SD= 12.7, $p < .05$). Low-efficacy teachers allocated nearly half of the observed time--and average of 214.5 minutes of 234 total minutes observed, or 48% (SD= 79.4, $p < .05$). Students of low-efficacy teachers spent more time off task when the teacher was focusing on small groups, and high-efficacy teachers maintained a higher level of student engagement over all the students. High-efficacy teachers allocated double the amount of time as low-efficacy teachers to whole-class instruction (an average of 65.5 minutes compared to an average of 30.0

minutes, respectively. The result may be that students spend more time off task and in transition when working alone, and greater levels of adult supervision occur in whole-group instruction.

Feedback patterns. Gibson and Dembo's (1984) study suggested that expectations that are more global, such as those in the construct of teacher efficacy, might influence feedback behaviors and teacher persistence. Teachers with higher levels of confidence in their ability to teach, and who believe students can learn, may communicate higher expectations to students that result in higher achievement, and they display less criticism and greater persistence in helping students answer correctly before going on to another student or another question. This possible connection between teacher efficacy to student achievement provided one of the guiding principles for the research being conducted for this dissertation, especially since student achievement is so closely associated with standards-based curriculum and assessment.

The Gibson and Dembo (1984) instrument, the Teacher Efficacy Scale, influenced nearly every study of teacher efficacy following its development. It was common to make minor modifications to the items, but major changes might have invalidated the resulting data, so the instrument was used in virtually the same form. By 1999, the Teacher Efficacy Scale (TES), the measure of teacher efficacy used across various studies, had been revised by Woolfolk and Hoy in 1990, and was updated again by Tschannen-Moran and Woolfolk Hoy in 2001.

Expansion of the Teacher Efficacy Scale

As research studies in teacher efficacy matured, closer examinations began between efficacy and concepts such as teacher expectations, teacher education programs, and classroom management skills. Woolfolk and Hoy (1990) reasoned that, by measuring preservice teachers' sense of efficacy, they could clarify the concept of efficacy itself. They enlisted 182 students enrolled in a teacher preparation program--155 women and 27 men. Nearly all (87%) were second- and third-year students in a four-year program; the remaining 13% were equally divided between first- and fourth-year students. Woolfolk and Hoy chose to use a version of Gibson and

Dembo's (1984) Teacher Efficacy Scale (TES). They revised it by (a) selecting only those items that resulted in "acceptable reliability coefficients" (Woolfolk & Hoy, p. 85) and (b) those items that referred to the adequacy of the teacher preservice program. Specific values of what constituted acceptable reliability quotients were not included in the report of the study. The result was 20 items that met the criteria--16 and 4, respectively. Woolfolk and Hoy added the two original items from the Rand study, bringing the total to 22 items. Responses were marked on a Likert scale of 6 points--from "strongly agree" to "strongly disagree." These 22 items would measure what Woolfolk and Hoy called "the predictor variables of general teaching efficacy and personal teaching efficacy" (p. 85).

The following sample items were from the 22-item *Teacher Efficacy Scale*, developed by Woolfolk and Hoy (1990). The first two samples loaded on Factor 1 (Personal Efficacy), and the second pair loaded on Factor 2 (Teaching Efficacy).

5. My teacher training program and/or experience has given me the necessary skills to be an effective teacher.
6. If a student in my class becomes disruptive and noisy, I feel assured that I know some techniques to redirect him/her quickly.
7. A teacher is very limited in what he/she can achieve because a student's home environment is a large influence on his/her achievement.
8. If students aren't disciplined at home, they aren't likely to accept any discipline (p. 83).

Woolfolk and Hoy (1990) compared preservice teacher sense of efficacy to three aspects of teacher attitudes about control and behavior in the classroom. The three aspects of control examined and compared in the study were pupil control ideology, motivational orientation, and bureaucratic orientation. Woolfolk and Hoy described pupil control ideology as existing along a continuum from custodial (regimentation, punishment, watchful mistrust, and a rigid and highly controlled environment) to humanistic (an educational community where students learn through cooperative interactions and experience). They also reported that teachers whose orientation is custodial tended more toward external locus of control, were more authoritarian in their belief system, and were less progressive in their educational attitudes. Woolfolk and Hoy connected to the purpose of their study by relating low-efficacy teachers with a stronger motivational orientation to custodial management. They described the aspect of motivational orientation as being a function of the level of control a teacher displays. For example, if the events in the classroom are inclined to be more informational than controlling, students tend to be more autonomous, more responsible, and more intrinsically motivated. The third aspect, bureaucratic orientation, relates to a person's commitment to the set of values, attitudes, and behaviors that are characteristic of a particular bureaucracy. Including bureaucratic orientation with pupil control ideology and motivational orientation extends the general idea of control in the school environment from student to school climate, with the teacher at the pivot point. Woolfolk and Hoy examined teacher sense of control over students as well as how the teacher perceives that the bureaucracy is controlling them.

Participants in the study (Woolfolk & Hoy, 1990) were 182 liberal arts majors enrolled in the teacher education program at a state university on the East coast. They were predominantly second- and third-year students (87%), and the remaining students (13%) were equally divided between first- and fourth-year students. Of the group of 182 preservice teachers, 104 were in the elementary education program, 78 were enrolled in the secondary education program.

The questionnaire distributed to the selected sample of preservice teachers was comprised of a set of four instruments and a request for certain demographic information. The first instrument was the modified Teacher Efficacy Scale (Gibson & Dembo, 1984) consisting of 22 items with a 6-point, Likert-type response scale ranging from strongly disagree to strongly agree and was aimed at measuring self-efficacy. The second instrument, for measuring Pupil Control Ideology, was developed by Willower, Eidell, and Hoy (1967) and consisted of 20 items with a Likert-type scale. No further description or explanation of the scale was provided except that

“… the higher the scores, the more custodial the respondent’s orientation toward pupil control” (p. 85). Motivational orientation, the third aspect, was measured by the instrument *Problems in School Inventory* developed by Deci, Schwartz, Sheinman, and Ryan. The instrument was made up of eight vignettes, each of which presented four possible solutions--highly controlling (HC), moderately controlling (MC), moderately autonomous (MA), and highly autonomous (HA). Responses were along a 7-point, Likert-type scale from very inappropriate (1) to very appropriate (7). The fourth instrument, the *Work Environment Preference Schedule* or WEPS developed by Gordon, was selected to measure bureaucratic orientation. It contained 24 items, each with a 5-point Likert-type scale ranging from strongly disagree to strongly agree. The four-instrument set made up what Woolfolk and Hoy (1990) labeled the questionnaire.

Woolfolk and Hoy (1990) distributed the questionnaires during regularly scheduled classes to six groups of preservice teachers who were participating in the study. Of the participants present to receive them, 88% returned usable forms. The analysis procedures started with a factor analysis--to confirm the existence of the two factors (Personal Efficacy and Teaching Efficacy) for the sample used. To determine the relationships between linear combinations of both dependent and independent variables, Woolfolk and Hoy used a canonical correlation. They then used a multiple regression to determine the correlation and interaction of both Personal Efficacy (PE) and Teaching Efficacy (TE) with pupil ideology, motivational orientation, and bureaucratic orientation. The predictor variables of PE and TE were entered with the interaction between them, coded as their simple cross product. Each predictor was then excluded in order to determine its unique contribution to the variable under question.

Woolfolk and Hoy (1990) completed two analytic procedures. They first replicated the two-factor analytic solution used by Gibson and Dembo (1984) and followed it with an analysis to determine if the data might yield more than two factors, replicating the work of Guskey in 1988. The two factors that emerged from Gibson and Dembo’s work were Personal Efficacy and

Teaching Efficacy. As in Gibson and Dembo's work, Woolfolk and Hoy focused on two aspects of personal efficacy: (a) responsibility for positive student outcomes and (b) responsibility for negative student outcomes. In the first statistical analysis of the data, teachers' responses to the TES were submitted to principal axis factoring, and both oblique and orthogonal rotations were used to compare item loadings and the degree of correlation between factors. Woolfolk and Hoy noted that the two factors were essentially uncorrelated (r = .008, with a delta value set at 0), and the pattern of factor loading was highly similar for both rotations. Because of the degree of similarity, Woolfolk and Hoy chose varimax rotation as the final solution. In Woolfolk and Hoy's study, the two factors that emerged from their study were Personal Efficacy and Teaching Efficacy, and they accounted for 27% of the variance.

In the second series of analyses, a principal components factor analysis using Kaiser's criterion of eigenvalues greater than 1 in combination with Catell's scree test resulted in the emergence of three factors that accounted for 32.8% of the variance. Item loadings and degree of correlations between factors were determined by both oblique and orthogonal rotations. The second round of analytic procedures yielded three factors. Teaching efficacy (Factor 2) remained nearly the same (Woolfolk & Hoy, 1990), but the factor of personal efficacy divided into two sub-factors that showed a moderate correlation. They reflected teacher sense of personal responsibility for student outcomes--both positive (Factor 3) and negative (Factor 1). Results indicated a moderate correlation between Factors 1 and 3 (r = -.402, with a delta value set at 0). Factor 2, general teaching efficacy, remained virtually the same in both the two-and three-factor solutions. Two moderately correlated factors emerged from the personal efficacy factor of the first analysis. Factor 1 reflected teacher sense of personal responsibility for negative student outcomes, and Factor 3 reflected teacher sense of personal responsibility for positive student outcomes. The results were consistent with the findings of Gibson and Dembo (1984) and with Guskey (1988).

Woolfolk and Hoy (1990) created efficacy scales by computing an unweighted average of the responses that loaded at .35 or higher on that factor. Twelve items and eight items met the criteria for PE and TE, respectively. Because there were unequal numbers of items for PE and TE, the final scale scores were standardized, and Cronbach's alpha was .74 for the TE scale and .82 for the PE scale.

Woolfolk and Hoy (1990) found that the teachers' scores on the PE and TE scales were uncorrelated ($r = .07$), and chose to compute the correlations for the two dimensions of personal and teaching efficacy separately. They also found that teaching efficacy was significantly correlated with pupil control ideology ($r = -.50$, $p < .01$); "The more the subjects believed in the power of the school to overcome home and background factors, the more humanistic their orientation toward pupil control" (p. 86). The findings for personal efficacy, however, were not significantly correlated with pupil control ideology ($r = -.04$). Teaching efficacy showed a negative relationship to bureaucratic orientation ($r = -.42$, $p < .01$); "The more the teachers believed in the power of the school in determining student achievement (teaching efficacy), the less they subscribed to a bureaucratic perspective" (p. 86). Personal efficacy also significantly correlated with bureaucratic orientation ($r = .18$, $p < .05$). For example, "teachers with low teaching efficacy are more bureaucratic than teachers with high teaching efficacy, but teachers with low personal efficacy have a less bureaucratic perspective than do those high in personal efficacy" (p. 88). Other relationships were also pointed out. For example, to be able to compare both PE and TE, Woolfolk and Hoy recoded the responses to the TE scores, so that the higher the score on both scales, the more efficacious the responses were. Subjects were categorized as either high or low on PE and TE based on the calculations of the median splits and the mean PCI (Pupil Control Ideology) and WEPS (Work Environment Preference Schedule) score in each category. Dividing teachers into the high-low categories of PE and TE enabled the researchers to interpret a significant interactive effect of the dimensions of efficacy on pupil control ideology and bureaucratic orientation.

One of the reasons Woolfolk and Hoy (1990) designed their study was to determine if the results of previous research could be replicated. For example, they based their analytic processes largely on those of the Gibson and Dembo study (1984). Some of the results of using similar analyses were consistent with the findings of Gibson and Dembo's work; however, some findings were unexpected. The relationships and correlations between (a) teaching efficacy and pupil control ideology; (b) personal efficacy and pupil control ideology; (c) teaching efficacy and bureaucratic orientation; and (d) personal efficacy and bureaucratic orientation were all expected to be similar, and they were. Two relationships were unexpected, however. First, teachers who had confidence in their personal abilities to influence students (high PE) tended to have a perspective that was slightly more bureaucratic. Second, neither dimension of teacher efficacy was significantly correlated to motivational orientation.

The interactive effect between high or low efficacy and both pupil control ideology and with bureaucratic orientation emerged with more detail in Woolfolk and Hoy's (1990) study than in previous research. The results seem confusing and somewhat inconsistent, but closer examination shows two competing formulations. In one instance, the findings are consistent with previous studies. In another instance, a secondary level of differentiation within personal efficacy provided evidence of new relationships. For example, personal efficacy and teaching efficacy, two factors discussed in Gibson and Dembo's (1984) TES and Ashton and Webb's (1986) work, were also strongly supported by the data in Woolfolk and Hoy's analyses. A second consistency was within the factor of teaching efficacy, although two different populations were involved. In studies spanning several years, experienced teachers held similar beliefs as preservice teachers about teaching in general (Wenner, 1993, 1995, 2001).

Information about new relationships came from the analyses of the teachers' beliefs about personal teaching efficacy. Differences within the data showed two related aspects: (a) responsibility for positive student outcomes, and (b) responsibility for negative student outcomes. Further differences were evident for prospective teachers. For example, prospective teachers with high teaching efficacy showed a greater tendency to be more humanistic in their pupil control ideology than prospective teachers with low teaching efficacy. That tendency, however, was limited to only those prospective teachers whose sense of personal efficacy was high--they believed in their ability to make a difference in student outcomes (Woolfolk & Hoy, 1990).

An unanticipated degree of complexity appeared in comparisons of canonical correlations and interactions. A positive relationship was indicated in the case of personal efficacy and bureaucratic orientation, and a negative relationship emerged between teaching efficacy and bureaucratic orientation. The canonical correlations, however, showed that a positive relationship between personal efficacy and bureaucratic orientation existed only under certain conditions. The conditions for the positive relationship, for example, are when teachers are more humanistic in their control orientation or when they are less optimistic about the power of education to overcome a poor home environment (Woolfolk & Hoy, 1990).

Woolfolk and Hoy (1990) identified the following components in their study: (a) two factors of the efficacy construct--personal efficacy and teaching efficacy; and (b) three aspects of control--pupil control ideology, motivational orientation, and bureaucratic orientation among prospective teachers. With at least six permutations and combinations possible from the components, and accessibility to a variety of new analytic processes, it was not completely unexpected that the discussion and recommendations would be complex. That complexity gave rise to several specific recommendations.

Recommendations from the expansion of the TES. The first, and perhaps most important, recommendation Woolfolk and Hoy (1990) made was to emphasize the necessity of researchers' establishing criteria that identify both high and low efficacy, despite how efficacy is conceptualized and measured. Determining cutoff points for high and low scores along a single continuum may be appropriate if the measure of efficacy is unidimensional; however, if the measure is multidimensional, a Likert-type item and scale allow varied analyses and findings and the use of composite scores. This results in a second recommendation. Woolfolk and Hoy strongly suggest that researchers look beyond composite scores to identify samples of high and low efficacy. For example, if the two dimensions, personal efficacy and general teaching efficacy, are used when designing items for the instrument, creating a composite score appears to nullify, or at least diminish, the results. Developing methods of measurement that show mixes and combinations of high and low scores on both/either efficacy scale might clarify both the construct and teachers' beliefs.

A third recommendation is based on the language of the items. Woolfolk and Hoy (1990) found that most of the items relating to teaching efficacy were stated in the negative, and the more efficacious response was disagreement with the item. Oppositionally, all the items related to personal efficacy were statements in the positive, and the more efficacious response was agreement. They suggest that this positive-negative orientation of the items may have influenced the direction of the response and recommend further research into the phenomenon.

The most crucial recommendation made by Woolfolk and Hoy (1990) involves the clear understanding of how efficacy is to be measured. A precise definition of efficacy for each study prevents efficacy from being mistaken for what might actually be "teachers' sense of political control within the school, their feeling of responsibility for student success or failures, sense of academic futility, general educational philosophy, belief in their power to influence students, or some composite of these beliefs" (p. 90).

Revised Validation of the Teacher Efficacy Scale

One of the recommendations from the work of Woolfolk and Hoy (1990) was that the aspect of the positive-negative orientation of the language used in the items of the Teacher Efficacy Scale should be examined more closely. Ten years passed after Gibson and Dembo's (1984) development of the Teacher Efficacy Scale (TES) before Guskey and Passaro (1994) examined the factor structure of efficacy research that had occurred over the previous two decades. Guskey and Passaro focused on the pair of questions from the Rand study (Armor et al., 1976), Gibson and Dembo's TES, and Woolfolk and Hoy's (1990) expansion of the TES.

Guskey and Passaro (1994) incorporated changes to the design of individual items into their study. They detected that all items loading on the personal efficacy factor used the referent *I,* were worded in a positive tone, and had internal locus of control (i.e., "I can"). Nearly all the items loading on the teaching efficacy factor used an "other" referent ("teachers"), a negative tone, and had external locus of control (i.e., "teachers cannot"). With this in mind, Guskey and Passaro modified the TES, administered the new form, analyzed the data, and compared the results to those of the original TES.

Guskey and Passaro (1994) identified 342 teachers for the study--283 inservice teachers, all of whom were from three medium-sized suburban/rural districts in two different states. There were 187 women and 96 men; all the teachers taught in grades K-12 and had an average of 10.4 years' teaching experience. Guskey and Passaro also included 59 preservice teachers--44 women and 15 men. All were students in a teacher preparation program at a large university in the West. All of the preservice teachers were in their junior or senior year of school, and all had completed several teaching practicum courses.

Guskey and Passaro (1994) used 16 items from Gibson and Dembo's (1984) original Teacher Efficacy Scale that yielded significant factor loadings--15 of which were also used by Woolfolk and Hoy (1990) in their extended study. Guskey and Passaro added 4 more items from

Woolfolk and Hoy to the 16 already selected, and included the two items from the Rand study (1976), bringing the total to 21 items. In the Rand study, the first question asked if the teacher felt "that 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" (p. 23). The second asked whether the teacher thought that "if I try really hard, I can get through to even the most difficult or unmotivated students" (p. 23).

The 16 items taken from Gibson and Dembo's (1984) instrument and the four taken from Woolfolk and Hoy (1990) loaded on two factors that were referred to as personal efficacy and teaching efficacy. Guskey and Passaro (1994), however, revised the items in two ways. The first was to create a balance between items that were worded to reflect the teachers' perceptions about themselves as individuals vs. their perceptions about teaching in general. The second was to create a balance between items worded with an orientation toward an internal locus of control vs. items worded with an orientation toward an external locus of control. Rather than refer to the second grouping as internal vs. external, Guskey and Passaro referred to them as positive (internal) and negative (external). To assist in tracking the item orientation, they created a 2 x 2 grid; marked the cells as P-I (personal-internal), T-I (teaching-internal), P-E (personal-external) and T-E (teaching-external); and reworded several items to maintain a four-way equilibrium.

Several of the items were then altered in the following ways: Of the total 21 items, 12 loaded principally on the personal efficacy dimension, and nine loaded on the teaching efficacy dimension. Seven of the 12 personal efficacy items reflecting a personal-internal orientation (P-I) were randomly selected and reworded to reflect either a teaching-internal (T-I) or personal-external (P-E) orientation. The Rand P-I item remained unchanged. Four of the nine teaching efficacy items, most of which had teaching-external orientation (T-E), were randomly selected and reworded to reflect either a personal-external (P-E) or teaching-internal (T-I) orientation. The Rand T-E item remained unchanged. All items were reassembled and numbered as in the

Woolfolk and Hoy (1990) study and used a 6-point Likert-type scale from "strongly disagree" to "strongly agree." The result was 21 items: five with a personal-internal (P-I) orientation, five with personal-external (P-E) orientation, five with teaching-internal (T-I) orientation and six with a teaching-external (T-E) orientation. The instrument with the revised items was administered to the inservice teachers at the beginning of a district-wide professional development meeting, and it took about 10-15 minutes to complete. Of the 283 inservice teachers participating, 92% of the submitted questionnaires were usable. The preservice teachers responded to the questionnaire at the beginning of regularly scheduled classes, and 95% of the returned questionnaires were usable.

In the report of their study, Guskey and Passaro (1994) stated that they were not attempting to identify a set of unknown factors but to explore the existing factor structure of teacher efficacy; thus, a principal components factor analysis was used to generate a 2-factor solution, which Guskey and Passaro called internal and external. Two factors were extracted based on Catell's scree test as well as agreement with factors identified in previous research. The item loadings and degree of correlations between factors were compared using both oblique and orthological rotations. With the delta value set at zero, the oblique rotation revealed that the two factors were only moderately correlated (r= -.23), suggesting that the two factors represent related, but relatively independent, constructs.

One of the tasks in the analysis was to check for inconsistencies in item responses on the altered scale across the four subscales--personal-internal orientation (P-I), personal-external (P-E), teaching-internal (T-I) orientation, and teaching-external orientation (T-E). The variance/covariance matrices and mean item responses of the experienced teachers were very similar. When the variance/covariance matrix of items responses of experienced teachers was compared to the data from preservice teachers using the DISCRIM Procedure from the Statistical Analysis System, no statistically significant differences were identified. The preservice teachers appeared to be more efficacious, but the results were not statistically significant. Numbers were not reported for nonsignificant results.

The final solution chosen was a varimax rotation with Kaiser Normalization (the orthological solution); this procedure attempts to minimize the number of variables that have high loadings on a factor. The varimax rotation converged in three interactions, yielding a 2-factor model that accounted for 32% of the total variance in item responses. The 11 items that were negative and external in orientation all loaded primarily on Factor 1--both personal and teaching items. Nine of the 11 items had factor loadings ≥ .40. The 10 items oriented as positive and internal all loaded primarily on Factor 2; nine with factor weights ≥ .40. Externally oriented items loading on Factor 1 with low scores indicated responses that were more efficacious, and internally oriented items loading on Factor 2 with high scores indicated the same. There was no evidence to support a personal versus teaching efficacy distinction. The reverse in loading sign between Factor 1 and Factor 2 was attributed to the difference in scale direction. As a result, *personal efficacy* and *teaching efficacy* were not considered separate factors. Guskey and Passaro (1994) found more statistical difference between beliefs about internal and external control than between personal and teaching efficacy beliefs.

Investigating the Factor Structure of the Teacher Efficacy Scale

Deemer and Minke (1999) set out to re-examine the factor structure of the Teacher Efficacy Scale and to assess the Woolfolk and Hoy hypothesis that the items on the TES cluster into factors, in part, because of the differences in item orientation across the two factors. To do the analysis they anticipated, Deemer and Minke (1999) began with an in-depth review of several efficacy measurement instruments. Their intent in doing the review was to find commonalities or differences between the factors that emerged from the analyses of the data in the studies in order to determine if teacher efficacy is a multi-dimensional construct composed of two relatively independent dimensions--personal efficacy and teaching efficacy.

With this background analyzed and discussed, Deemer and Minke (1999) undertook the task of attempting to assess Woolfolk and Hoy's (1990) hypothesis of bias being inherent in the positive-negative language orientation of the items by establishing a matrix of sorts that would examine the effect of positive versus negative orientation in item wording across both the internal and external dimensions. The four combinations were internal positive, internal negative, external positive, and external negative. Their hypothesis was that if Guskey and Passaro's (1994) analysis was accurate, then the items would load the same way on the internal-external factors regardless of the orientation of the wording. If, however, it was not so, then a different factor structure would emerge. It might still be a two-factor structure, but the two factors would be positive-negative rather than internal-external. Another possibility was a four-factor structure in which the items clustered into independent positive and negative dimensions of both personal-internal and general-external items. (The designations of personal-internal and general-external were retained from the Guskey and Passaro study. In their study, Guskey and Passaro reported that all items with positive orientation used the referent "I," and all items with negative orientation used the referent "teachers." The "I can" items were then called personal-internal, and the "teachers cannot" items were called general-external.)

Further investigation of the 2-factor structure used by Gibson and Dembo (1984) and Guskey and Passaro revealed confusion about which dimensions of efficacy were being measured by the items in the instrument. For example, Gibson and Dembo explained their results in terms of outcome and efficacy expectations, citing Bandura (1973, 1977). Guskey and Passaro used the terms "personal-internal" and "general-external" to describe the dimensions that emerged in their study. Deemer and Minke (1999) argued that the items used as examples of the two factors in both previous studies did not match Bandura's definitions of both efficacy and outcome expectations as they claimed.

The participants in the Deemer and Minke (1999) study were 196 inservice teachers - the majority of whom were women - enrolled in summer graduate classes at a moderately sized university located in the northern United States. The teachers completed one of two forms of a modified Teacher Efficacy Scale (TES) and provided demographic information related to gender, ethnicity, highest level of education, grade level teaching assignment, age, and years of teaching experience. Their teaching experience ranged from 1 to 28 years, with an average teaching experience of 6.8 years for teachers who responded on Form A and 8.4 years for those who completed Form B. The teachers' grade levels ranged from Pre-Kindergarten to Grade 12, with the majority assigned to grades 1-12.

Deemer and Minke (1999) developed two equivalent forms of the instrument with all items written with an "I" referent except one; that item retained the referent "teachers." Each item, now changed to the first person, was then written in both positive and negative versions. This allowed the examination of the effect of only the positive and negative wording orientation in the factor structure. Each form of the modified TES--Form A and Form B--contained equal numbers of positive and negative items.

Form A was composed of 17 items: (a) five original personal-internal ("I can"); (b) four revised personal-internal ("I cannot"); (c) four original general-external; and (d) revised general-external. Form B also contained 17 items: (a) four original personal-internal; (b) five revised personal-internal; (c) four original general-external; and (d) four revised general-external. See Table 2.2 for examples of the items. No list of specific items on Form A and Form B was designated.

TES Items	
Original Form from Gibson and Dembo (1984)	Revised TES items from Gibson and Dembo (1984)
When a student does better than usual, many times it is because I exerted a little extra effort.	When a student does better than usual, it is often not because I exerted a little extra effort.
The influences of a student's home experience can be overcome by my teaching.	The influences of a student's home experience cannot be overcome by my teaching.
If a student masters a new concept quickly, this might be because I knew the necessary steps in teaching that concept.	If a student masters a new concept quickly, is not likely to be because I knew the necessary steps in teaching that concept.
The time spent in my class has little influence on students compared to the influence of their home environment.	The time spent in my class has a big influence on students compared to the influence of their home environment.
When a student is having difficulty with an assignment, I am usually able to adjust it to her/his level.	When a student is having difficulty with an assignment, I often have trouble adjusting it to her/his level.
I am a powerful influence on student achievement when all factors are considered (Woolfolk & Hoy, 1990).	I am not a very powerful influence on student achievement when all factors are considered.

Table 2.2: Sample items from Deemer and Minke (1999), original and revised.

Deemer and Minke (1999) submitted each form of the instrument to principal components analysis so that the effects of wording on the factor could be examined and to maintain consistency with previous research so comparisons could be made with Gibson and Dembo's (1984) findings as well as the findings of Guskey and Passaro in 1994. The earlier study also used principal components analysis, but Deemer and Minke's intention was to apply the same strategy as was reported earlier for estimating communalities because both techniques often yield similar results. As an additional strategy, Deemer and Minke also submitted the revised TES to a principal components analysis, and in the analyses of both Forms A and B, they followed similar procedures. They used two rotations using statistical analysis software--an oblimin rotation and an orthogonal rotation, and for both rotations, they found much similarity in the factor loadings. A confounding assumption was that they, as well as other researchers, believe that the construct of efficacy is multi-dimensional and correlated, which implies that correlation among factors should not be restricted. The outcome was that the oblimin rotation was used in reporting the factor analyses. Items were retained based on how the factor loadings clustered, and an assumption was made that items with factor loadings that differed significantly from the loadings of other items that clustered would not provide the correlations the researchers sought.

An examination of the four factors on each form showed that principal axis factoring did not adequately represent the data. Some item loadings were substantial, but the items did not cluster into meaningful constructs, and the factors did not make theoretical sense. For example, on Form A, two of the factors approximated the internal-external dimension, but the remaining two factors were defined by only one or two items, which made them difficult to interpret. On Form B, two factors were defined by single items, and the remaining factors contained two and three items; neither internal-external nor positive-negative distinctions were apparent. All four factors taken together accounted for 37.5% and 29.7% of the variance on Forms A and B, respectively. However, most of the variance on each form was accounted for by the first factor -- 21.6% on Form A and 13.7% on Form B. After an inspection of the scree plot for this analysis, the researchers concluded that they needed to investigate other models to explain the revised TES besides the four-factor model (Deemer & Minke, 1999).

The next question to be investigated was if the revised TES contained factors reflective of two dimensions, i.e., either personal-internal versus general-external or positive versus negative wording. To make a determination, a principal components analysis was completed, specifying two factors for each form. It showed that two factors did not adequately represent the data. Seven items that were a mix of the four types--personal-internal, general-external, positively worded, and negatively worded items--defined the first factor on Form A. Only one item unambiguously loaded on the second factor on Form A. Twenty-six-and-one-tenth percent of the variance was accounted for by the two factors; however, 21.1% of the 26.1% of the variance was related to the first factor. On Form B, a mix of eight items of the four types--personal-internal, general-external, positively worded, and negatively worded items -- defined the first factor and accounted for 12.9% of variance. The second factor was composed of two items and only 5.5% of variance. An examination of the scree plots for each form supported a one-factor model.

A principal components analysis specifying one factor was then completed for each form. Nine items on Form A best defined a single factor and accounted for 20.8% of the variance. On Form B, eight items explained 12.9% of variance. The authors state that the variance accounted for on both forms was low, but that it was consistent with the amount of variance accounted for in previous studies by Gibson and Dembo (1984) and Guskey and Passaro (1994). The items retained on both forms contained predominantly personal-internal items with both positive wording of "I can" and negative wording of "I cannot." On Form A, seven of the nine retained items were personal-internal. Participants' scores on that factor showed an internal consistency of _ = .81. On Form B, five of the eight were personal-internal and participants' scores showed lower internal consistency of _ = .66.

When factors and items were compared across Forms A and B, six items (five personal--internal; one general--external) from Form A that comprised the factor, also were found in the factor of Form B. Deemer and Minke (1999) describe them as modifications of the same original items from Gibson and Dembo (1984). The interpretation was that those items were comparable across forms, that is, their clarity was maintained and was not affected by positive versus negative wording. From these findings, Deemer and Minke (1999) state that ". . . the internal dimension of teacher efficacy discussed by Guskey and Passaro (1994) seems to have primarily defined the single factor on each form" (p. 8).

With the revision of the items to reflect both positive and negative orientations across internal and external influences on teaching, Deemer and Minke (1999) suggest that teacher efficacy, as measured by the TES, is actually unidimensional. When the instrument was redesigned to eliminate the confounding of wording bias, the items appeared to tap an internal dimension similar to the personal efficacy factor proposed by Gibson and Dembo (1984) and to outcome expectations defined by Bandura (1977, 1997). Most of the items referring to outside influences on teaching were deleted when the positive-negative wording bias was eliminated, and a separate external factor was not identified. The authors also state that the degree of overlap between Forms A and B further supports the conclusion that the wording problem has confounded prior interpretations of the TES. Despite varying the item orientation, both forms contained many of the same items; therefore, the two-factor structure that has been replicated throughout the literature appears to be at least partially an artifact of item wording and not the result of underlying, distinct construct dimensions.

A caution expressed by researchers Deemer and Minke (1999) is that the findings of the study should not be construed as "proof" that efficacy is a single-dimension construct. They do suggest that teacher efficacy may be more differentiated than the TES adequately captures. Although various studies have documented a relationship between levels of efficacy and instructional effectiveness, those relationships are weak because of the omnibus-type of instruments used (Pajares, 1997). Global instruments such as the TES present efficacy judgments without a context for interpretation and lack close relationships to specific teaching tasks, which might limit their predictive capabilities. Another perspective is that teachers' sense of efficacy may vary across the many tasks of teaching, so "… instruments that separately assess teachers' perceptions in specific domains of teaching can be expected to tap the variations in efficacy judgments and increase the predictive power of efficacy perceptions" (Deemer & Minke, p. 9). Studies that have included efficacy items reflective of particular classroom tasks and situations have documented greater dimensionality in teacher efficacy than originally proposed by Gibson and Dembo (1984).

Recommendations from the investigation of the factor structure. Deemer and Minke (1999) concluded that if the TES factors did not illuminate efficacy versus outcome expectations, as suggested by Gibson and Dembo (1984), or separate internal and external dimensions, as suggested by Guskey and Passaro (1994), then a re--examination of studies using the TES is warranted. Another conclusion was that the TES did not measure two clearly identifiable dimensions of efficacy as originally proposed--either personal efficacy and teaching efficacy or positive and negative. Deemer and Minke caution, however, that this conclusion does not imply that efficacy is a unidimensional construct. Instead, it implies that the TES did not measure two dimensions. This is in complete coherence with Woolfolk and Hoy's (1990) recommendation that researchers should determine, with as much specificity as possible, precisely what aspect of efficacy they are attempting to measure.

Deemer and Minke (1999) suggest that, to overcome difficulties inherent in interpretation of specifics when using an instrument with a more global design, future studies should devise measures that are more specific to teaching duties, e.g., classroom management, relating to parents, lesson planning, and working with colleagues. They also suggest that additional studies are needed to explore teachers' perceptions regarding external influences on teaching. The support for this suggestion comes from the diverse nature of outside influences on teaching beyond the role of parents that is not represented on the TES, influences such as the community in which the student lives and where the school is located, the student's peers, and social institutions to which the student or the student's family belongs.

Ohio State Teacher Efficacy Scale (OSTES)

A statement that appeared in previous studies (Deemer & Minke 1999, Woolfolk & Hoy 1990) declared that teacher efficacy is a construct not easily measured. Tschannen-Moran and

Woolfolk Hoy (2001) agreed and set out to develop a new instrument to withstand greater scrutiny. They attempted to address questions of reliability and validity that arose regarding existing measures and examine the commonplace two-factor structure, that is, internal/external, positive/negative, or efficacy/outcome expectation because the clear identification and meaning of the two factors has caused confusion and debate. Another major disagreement they addressed is the concern surrounding the conceptualization of teacher efficacy. For example, one of the major questions involves the extent to which teacher efficacy is specific to given contexts. Another question is related to the extent of efficacy beliefs being transferable across contexts. A third difficulty is ascertaining the appropriate level of specificity to be used when designing items for an instrument. Following an examination of each of the studies selected by Tschannen-Moran and Woolfolk Hoy, they attempted to address the issues through the design of a new measure of teacher efficacy.

The purpose of the paper, as described by Tschannen-Moran and Woolfolk Hoy (2001), was "… to explore issues related to the measurement of teacher efficacy and to propose a new measure" (p. 784) based on a model developed by Tschannen-Moran, Woolfolk Hoy, and Hoy. The model to which they refer is the product from an earlier study that examined the sense of efficacy and beliefs about control held by prospective teachers (Woolfolk & Hoy, 1990). Woolfolk and Hoy found two independent dimensions of efficacy, personal efficacy and teaching efficacy. Personal efficacy relates to the beliefs a teacher holds about her/his own abilities to teach; teaching efficacy is the set of beliefs about teaching in general and what teachers might be able to accomplish. They also found that personal efficacy could be subdivided into two other aspects--one of responsibility for positive student outcomes and the other of responsibility for negative student outcomes. The two subdivisions of personal efficacy did not provide strong enough evidence for Woolfolk and Hoy to justify keeping them as separate dimensions, so in the final analysis, all analyses were based on the pair of original, independent dimensions of personal efficacy and of teaching efficacy.

In their efforts to design a more reliable and valid measure of teacher efficacy, Tschannen-Moran and Woolfolk Hoy (2001) reviewed two types of instruments. Previous measures related to locus of control (internal vs. external) that were grounded in Rotter's (1966) Social Learning Theory, such as the Rand study (Armor et al., 1976) and those by Rose and Medway (1981) were the first type. Measures of teacher efficacy based on Bandura's (1977) social cognitive theory were the second type. The comparison between instruments based on Rotter's locus of control construct and Bandura's conceptualization of self-efficacy was the new lens through which Tschannen-Moran and Woolfolk Hoy would examine an updated instrument designed to measure teacher efficacy beliefs. They wanted an instrument that purported to measure teacher sense of efficacy for student engagement, efficacy for instructional strategies, and efficacy for classroom management. The construction of the new instrument took other researchers' recommendations, such as the balance of positive-negative orientation of the language of the items, into account.

One of the most important considerations of the new design was the attribution of the two-factor structure to Bandura's (1977) social cognitive theory and self-efficacy. For example, when a factor analysis yielded a two-factor structure, Gibson and Dembo (1984) matched them to the two expectancies of Bandura's social cognitive theory--self-efficacy and outcome expectancy. Gibson and Dembo labeled the first factor *personal efficacy* because they believed it reflected self-efficacy, and they labeled the second factor *general teaching efficacy*, because they believed it was related to outcome expectancy. Tschannen-Moran and Woolfolk Hoy (2001) reported that support continued for the two-factor structure, but as research continued with the Gibson and Dembo items, inconsistencies became evident.

Other studies (Deemer & Minke, 1999; Guskey & Passaro, 1994; Woolfolk & Hoy, 1990) also found a two-factor structure, but there was disagreement about what those two factors were. Researchers continued to use the TES instrument, but they were cautioned to conduct their own factor analyses because of inconsistencies. Also, it was suggested that they not rely exclusively on Gibson and Dembo's (1984) work. Tschannen-Moran and Woolfolk Hoy (2001) also suggested that the lack of clarity about the meaning of the two factors and because the factor structure is unstable, continued use of the Gibson and Dembo instrument is problematic.

Tschannen-Moran and Woolfolk Hoy (2001) reported facing a challenge in determining a level of generality or specificity in the items of their new measure of efficacy. The decision to be made was whether to develop items with language reflecting greater specificity of context that they expected would generate results that were more significant, or to aim for greater predictive value and generalizability that would come from items written with language that was more global. As examples of three different contextual areas, Tschannen--Moran and Woolfolk Hoy investigated and discussed classroom management, special education, and science teaching.

Tschannen-Moran and Woolfolk Hoy (2001) collected evidence to support the idea that the measures of teacher efficacy are open to large differences in interpretation. They presented the evidence in support of their rationale for developing a new instrument that addresses each of the issues and shortcomings of previous instruments. For example, the first discussion describes the findings of researchers who administered a set of measures to a sample of elementary and secondary teachers and found that a correlation of 0.64 existed between the Rand items and Gibson and Dembo's (1984) Teacher Efficacy Scale. A second finding was that the Teacher Efficacy Scale and the Teacher Locus of Control Scale (Rose & Medway, 1981) correlated at 0.47. A third correlation of 0.57 was found between the Teacher Efficacy Scale and Responsibility for Student Achievement Questionnaire developed by Guskey. Tschannen-Moran and Woolfolk Hoy point out that the data show that the constructs are related but an imperfect overlap exists, and the question becomes one of how much of the analysis supports teacher efficacy and how much supports something else.

Tschannen-Moran and Woolfolk Hoy (2001) involved 10 participants of a university seminar (two researchers and eight graduate students) to review previous teacher efficacy scales. Bandura (1999) believed that intervals between points were too wide on 5- or 7-point Likert scales, and that they were insufficient to convey any subtleties associated with teacher beliefs. He advocated consideration of a 100--point Likert scale, but the most widely-used instruments include a 5- or 7-point scale. One of the characteristics in Bandura' instrument criticized by the seminar participants was that the items across seven subscales in his instrument did not accurately represent the types of activities that were a regular part of a teacher's classroom responsibilities. They used a process of nominate, discuss, and revise to settle on 52 items--some that came from Gibson and Dembo's (1984) instrument, some from Bandura's (1997) instrument, and some that were designed especially for the new instrument. The characteristics they selected were efficacy for student engagement, efficacy for instructional strategies, and efficacy for classroom management. The new measure, called the Ohio State Teacher Efficacy Scale or OSTES*,* was examined and refined in three successive studies.

Tschannen-Moran and Woolfolk Hoy's (2001) first study involved 224 teachers. The participants represented diversity across several dimensions: (a) preservice or experienced, (b) years of teaching experience, (c) age, (d) ethnicity, and (e) gender. This level of diversity among the first participants seemed to contribute to the credibility of the instrument among subsequent classroom teachers involved in instrument development.

The instrument developed by Tschannen-Moran and Woolfolk Hoy (2001) was comprised of 52 items with a 9-point scale ranging from "1-nothing, 3-very little, 5-some influence, 7-quite a bit, to 9- great deal" (p. 796). A 4-point scale was also included for respondents' rating of the level of importance of each item relative to effective teaching, e.g., 1-

not at all, 2-somewhat, 3-important, 4-critical. When the data was analyzed, it was noted that all teachers reported the tasks as "important" to "critical," and no items were eliminated based on importance ratings.

The teachers' responses to the 52 items were submitted to principal-axis factoring with a varimax rotation, and ten factors emerged with eigenvalues greater than one, accounting for 57.2% of the variance in the respondents' scores. When rotation failed to converge after 25 iterations, the unrotated factor matrix was examined, and a criterion for factor loadings of 0.60 was set for selecting items for further analysis. As a result, 31 items emerged with loadings ranging from 0.62 to 0.78. One item pertaining to the area of motivation with a loading of 0.60 was also included, creating an instrument with 32 items. The decision to include that specific item was based on three characteristics of that particular item. First, its loading value was within 0.05 of the cut-off, and second, the item referred to motivation. A seminar group of teachers believed that motivation was a critical task of teaching. Third, motivation was not well represented in the item pool. This new set of 32 items became the measure of teacher efficacy that would be used in the second study (Tschannen-Moran & Woolfolk Hoy, 2001). In this phase of the study, unfortunately, the authors did not provide an explanation of the two factors, so the reader is left to wonder whether the first factor that emerged from the data analysis was the more general teaching efficacy or the more specific personal efficacy.

The second study (Tschannen-Moran and Woolfolk Hoy, 2001) in the process of developing a new instrument involved 217 teachers--70 preservice and 147 inservice. The participants in the second study again represented diversity across several dimensions: (a) preservice or experienced teachers, (b) years of teaching experience, (c) age, (d) ethnicity, and (e) gender. This level of diversity among the first participants continued to contribute to the credibility of the instrument among subsequent classroom teachers involved in the study.

In the course of Study 2, the discovery of some weakness of the management factor led to the third study and the reduction of the instrument to 18 items from the previous set of 32. The analysis of the 32 items using principal-axis factoring with varimax rotation resulted in eight factors with eigenvalues greater than one (Tschannen-Moran & Woolfolk Hoy, 2001). The eight factors accounted for 63% of the variance in the respondents' scores. The authors then examined both solutions--a two-factor and a three-factor--suggested from the results of a scree test. Items related to classroom management loaded across both factors almost equally in the two-factor solution, but the loadings were low. However, management emerged as a separate factor in the three-factor solution, and the other two factors were more clearly specified. Because of the importance of classroom management to effective teaching, the authors retained the three-factor solution, because it better represented the tasks of teaching. They decided to use the three-factor solution to identify other items that could be deleted from the instrument without reducing its reliability or validity (2001), which were _ = 0.82 for efficacy for engagement, _ = 0.81 for instruction, and _ = 0.72 for management.

Decisions on the removal of items were based on the following three criteria: (a) items that had the lowest loadings on each of the three factors, (b) items that clearly loaded on more than one factor, and (c), items that seemed redundant. The same three subscales, (a) efficacy for student engagement, (b) efficacy for instructional strategies, and (c) efficacy for classroom management were maintained (Tschannen-Moran & Woolfolk Hoy, 2001). The two following items were provided as examples.

> "How much can you do to adjust your lessons to the proper level for individual students?" and "To what extent are you able to tailor your lessons to the academic level of your students?" both loaded on the same factor and were moderately correlated (r = 0.54). Therefore, we eliminated the second item because it had a lower loading on the factor. (p. 797)

The result was an 18-item instrument. Eight items represented efficacy for student engagement, seven items represented efficacy for instructional strategies, and three items represented efficacy for classroom management. When a varimax rotation of the 18 items in the respondents' scores was completed, the three factors accounted for 51% of the variance. Reliabilities for the subscales were computed at 0.82 for efficacy for *engagement*, 0.81 for efficacy for *instruction*, and 0.72 for efficacy for *management* (Tschannen-Moran & Woolfolk Hoy, 2001).

A process Tschannen-Moran and Woolfolk Hoy (2001) called "second-order factor analysis" was completed using the data acquired from combining Study 1 and Study 2 rather than maintaining complete separateness. A principal-axis factoring of three subscales revealed one strong factor with loadings ranging from 0.74 to 0.84.

> The emergence of this second-order factor and the moderate positive correlations of the three subscales suggested that the 18 items could be considered to measure the underlying construct of efficacy and that a total score as well as three subscale scores could be calculated based on the 18 items. (p. 798)

Another principal-axis factor analysis specifying one factor indicated that all 18 items loaded on the single factor with values ranging from 0.48 to 0.70. The calculated reliability for the 18-item scale was 0.95.

Determining the level of construct validity of the new instrument was another goal, so the researchers assessed its correlation to other measures. The instrument under scrutiny was the OSTES (Tschannen-Moran & Woolfolk Hoy, 2001) and it was compared to the two items from the Rand study (Armor et al., 1976), Hoy and Woolfolk's (1993) 10-item adaptation of Gibson and Dembo's Teacher Efficacy Scale (1984), the Pupil Control Ideology Form (Willower, Eidell, & Hoy, 1967), and a Work Alienation Scale developed by Forsyth and Hoy. The results indicated that the total scores on OSTES positively related to each Rand item (r = 0.35 and 0.28 for Item 1 and Item 2, respectively, $p<0.01$) and both factors of the Gibson and Dembo measure (personal teaching efficacy: $r = 0.48$, $p<0.01$; general teaching efficacy: $r = 0.30$, $p<0.01$).

Discriminant validity was measured using a survey of work alienation because alienation was "defined in terms of the extent to which individuals fail to experience intrinsic pride or meaning in their work" (Tschannen-Moran & Woolfolk Hoy, 2001, p. 798) and because of its being conceptually opposite to teacher efficacy. The results indicated that teacher efficacy was negatively correlated to work alienation: $r = -0.31$, $p<0.01$. In other words, respondents who scored high on the work alienation scale were expected to show low scores on the teacher efficacy scale, specifically OSTES, and vice versa.

A measure of pupil control ideology was also administered to the teachers. This instrument was included because it claimed to measure **"**the extent to which a teacher takes a custodial rather than a humanistic stance toward students and has been related to teacher efficacy as measured by the Gibson and Dembo instrument" (Tschannen-Moran & Woolfolk Hoy, 2001, p. 798). This instrument was also selected because of its philosophical opposition to a teacher sense of efficacy. The results indicated that teacher efficacy, as measured by OTES, is negatively correlated to pupil control ideology, i.e., teachers with a strong sense of their own efficacy tended to be less custodial in their attitudes toward students: $r = -0.25$, $p<0.01$. In an additional consideration, correlations were run again without the scores of the preservice teachers to insure that the data were not skewed by the inclusion of preservice teachers. The results are reported as being very similar, but no specific data were included to support the claim.

The findings from Study 2 were encouraging, and the factors were conceptually sound representations of the three teaching tasks, but a weakness of the instrument was in the management items. A third study was considered to improve the management items and enhance the strengths of the new instrument.

In the third study, 410 teachers participated--103 preservice teachers, 255 inservice teachers, and 52 who did not indicate their teaching status. In each phase of the overall study, participants were asked to provide demographic information. The characteristics of age, years of teaching experience, gender, and ethnicity were used in the two previous studies and were used again to maintain continuity and consistency.

As in the two previous stages of Tschannen-Moran and Woolfolk Hoy's (2001) study, the purpose of the third study was to further refine the teacher efficacy instrument. Tschannen-Moran and Woolfolk Hoy received affirmation from other researchers' findings related to the factors of engagement and instruction, but it was recommended that the items related to management be eliminated. There was resistance to dropping the items, because Tschannen-Moran and Woolfolk Hoy suspected that since there were only three items related to management, the fault might be in the brevity of the scale. Another influence on the decision to keep the third factor was that they wanted to capture a wider range of teaching tasks. Both the Rand items (Armor et al., 1976) and the items from the Gibson and Dembo (1984) instrument focused on negative aspects of teaching.

In addition, both preservice and inservice teachers agreed on the importance of management in the context of teaching and learning. To accomplish the refinement, 19 students in a class at a Midwestern university that included 17 teachers and 2 teacher educators, provided feedback on several items designed to improve the management subscale. Using a principal-axis factoring with varimax rotation of the 36 items, the solution yielded four factors with eigenvalues greater than one, and they accounted for 58% of the variance in the scores. A scree test revealed three factors that were the same as the three identified in the second study--12 items related to efficacy for student engagement; 15 items related to efficacy for instructional strategies; and 9 items related to efficacy for classroom management (Tschannen-Moran & Woolfolk Hoy, 2001). The revised instrument was administered to the teachers and the data collected in preparation for the analyses.

The researchers (Tschannen-Moran & Woolfolk Hoy, 2001) wanted to shorten the length of the instrument without sacrificing quality, so they reduced the scale by extracting the eight items in each factor with the highest loadings. They used principal-axis factoring with varimax rotation again on the 24 items, and it yielded the same three factors--efficacy for student engagement, efficacy for instructional strategies, and efficacy for classroom management. An efficacy subscale was also calculated for each factor by finding the mean of the responses to the eight items that loaded highest on that factor. Reliabilities for the subscales were: (a) efficacy for student engagement--0.87, (b) efficacy for instructional strategies--0.91, and (c) efficacy for classroom management--0.90. Intercorrelations between the three subscales for efficacy for instruction, efficacy for management, and efficacy for engagement were 0.60, 0.70, and 0.58 respectively ($p < 0.001$).

Following the strong results of the 24-item subscale analysis, the researchers (Tschannen-Moran & Woolfolk Hoy, 2001) decided to reduce the number of items even further. The top four items with the highest loadings on each scale produced similar results to the 24-item analysis. The reliabilities for the12 items representing the three factors were 0.86 for efficacy for instruction, 0.86 for efficacy for management, and 0.81 for efficacy for engagement. Another strategy testing the reliability of the scale looked for intercorrelations between the long and short forms for the total scale and the three subscales. It resulted in a range from 0.95 to 0.98. Subsequently, the researchers were encouraged to continue with additional analyses.

With a concern that the overall factor structure might be different for the two groups of teachers, the researchers (Tschannen-Moran & Woolfolk Hoy, 2001) examined the factor structure for preservice teacher responses (n= 111) and inservice teacher responses (n= 255) separate from each other. The analysis of the inservice teacher responses showed the same three factors as in the second study, accounting for 54% of the variance in the long form (24 items) and 65% of the variance in the short form (12 items). The 24--item, long form included all 12 items from the short form. The factor structure for the preservice teachers was less distinct with three factors, which the researchers attributed to the minimal amount of actual teaching experience among those in the preservice group. The results of a principal-axis factoring called for a single factor (Un--named.) from the responses of the preservice teachers on both the 24-item and the 12-item subscales. All items' factor loadings ranged from 0.60 to 0.85 and accounted for 57% and 61% of the variance on the subscales, respectively.

The next step was to repeat a similar process with the responses of the inservice teachers (N= 255). Since this group of teachers had classroom experience, the researchers used principal-axis factoring on all the items for the three teacher efficacy subscales of *instruction, management,* and *engagement* (Tschannen-Moran & Woolfolk Hoy, 2001). The solution produced a single strong factor from the 24-item instrument, accounting for 75% of the variance. With the 12-item instrument, a single factor also emerged and accounted for 68% of the variance. The authors then concluded that both the 24- and 12-item instruments could be considered to measure teacher efficacy and that both the subscales and total scores could be calculated. An additional principal-axis factor analysis specifying one factor was next. All items loaded on the single factor within a range from 0.49 to 0.75 for the 24-item scale and from 0.49 to 0.76 for the 12-item scale. The reliabilities for the 12- and 24-item scales were 0.94 and 0.90, respectively.

In examining all the results, it seemed appropriate to the researchers (Tschannen-Moran & Woolfolk Hoy, 2001) to use the total scores for the preservice teachers because they have not yet had fulltime classroom responsibilities with the subscale categories. They made the decision to use the total score and a single factor for preservice teachers and to use a three-factor structure and the three subscales for inservice teachers.

Again, using second-order factor analysis, all data from the sample in Study 3 were analyzed using principal-axis factoring. That is all data relating to the 12--item form was included in one analysis, and all the data relating to the 24--item form was included in a separate analysis.

To determine construct validity of both forms of the OSTES (Short Form and Long Form), they were subjected to an assessment of their correlation to other measures, including the Rand items and Hoy and Woolfolk's (1993) 10-item adaptation of Gibson and Dembo's Teacher Efficacy Scale (Tschannen-Moran & Woolfolk Hoy, 2001). Total scores on the 24-item long form of the OSTES, with n=410, were positively correlated to both the Rand 1 (external locus of control) and Rand 2 (internal locus of control) items, $r = 0.18$ and 0.53, respectively, with $p < 0.01$. They also positively correlated to both factors of the Gibson and Dembo (1984) measure: personal teaching efficacy ($r = 0.64$, $p<0.01$); and general teaching efficacy ($r = 0.16$, with $p<0.01$). The researchers (Tschannen-Moran & Woolfolk Hoy) reported that correlations were similar with the short form. The data showed the strongest correlations between OSTES and other measures with scales that assess personal teaching efficacy. Evidence of a lower correlation emerged between OSTES and other measures in the dimension of general teaching efficacy.

In the second and third studies, data from the previous studies provided a basis of comparison for new data and generated new insights about efficacy. It was hoped that the recommendations arising from the results would affect teacher education programs and professional development for inservice teachers.

Tschannen-Moran and Woolfolk Hoy (2001) found that the results of the development of the OSTES mirror the findings of previous research. They suggested that the OSTES--as well as other measures--are less effective in capturing the essence of efficacy related to Bandura's second dimension, outcome expectancy, or, as Gibson and Dembo label it, general teaching efficacy. They also concluded that the OSTES went beyond other measures and addressed a wider range of teaching tasks, had a unified and stable factor structure, and assessed a range of capabilities that teachers consider important to good teaching.

Issues Related to Measuring Efficacy in Teacher Education

Throughout the development of instruments designed to measure teacher self--efficacy, several issues emerged. One is related to the philosophy of cognition and learning theory that permeates the instrument design; a second is related to various aspects of the methodology used in the studies, including the factor structure and the psychometrics of item design; and the third relates to the interpretations that emanate from the findings. Although the three threads are distinct, they are not discrete. Each issue influences the others, and in turn, is influenced by them and cannot be discussed without referring to the others. For example, in the report of Rose and Medway (1981), there is a comment that a sense of efficacy is "... a concept similar to locus of control" (Rose & Medway, p. 186). At the time, the development of the construct of self-efficacy was relatively new, and rigorous studies to describe it further had not yet been designed. Rose and Medway chose to develop items with language that parallels Bandura's ideas about tasks and specificity but continued to support the connection to Rotter's I-E scale (1966). Because of a naive understanding of Bandura's social learning theory and its "daughter-theory" of self-efficacy, Rose and Medway attribute their findings to one foundation (Bandura) when it should be another (Rotter). Further comparisons will be included in the chapter summary.

Measuring Efficacy Among Science Teachers

Precursors to Measuring Teacher Efficacy Among Science Teachers

One of the early statements from the article by Riggs and Enochs (1990) states that the dimension of Personal Teaching Efficacy (PTE), as described in the literature starting with Gibson and Dembo in 1984, differs from Bandura's original premise that self-efficacy and outcome expectancy are two distinct variables. On the contrary, Bandura's (1997) own explanations consistently refer to the relationship between the two dimensions as reciprocal causation. It is not a linear, cause-effect relationship--if I behave in this way, consistently and continuously, I expect the same result consistently and continuously. Rather, each dimension provides information in a feedback loop to the other dimension, which changes to accommodate the new information, which, in turn, effects a change in the subsequent expression of the first dimension, and so on continuously. For example, a beginning science teacher who has been successful in the academic environment may have a sense that s/he can write lesson plans that address the cognitive, affective, and psychomotor domains of her/his students. S/he uses a format consistent with a particular learning cycle, and is very familiar with the theoretical aspects of a classroom lesson and expects the students to be able to collect data for an ongoing project. However, in her/his first classroom teaching experiences, s/he realizes very quickly that her/his judgment about her/his ability to teach a lesson will be modified. Perhaps the lesson goes moderately smoothly, and the students complete the data collection s/he directed them to do. The next time the novice teacher develops an interactive lesson, s/he remembers what went well with the students and what should be revised. S/he uses reflective feedback from the first lessons to modify not only her/his judgment regarding her teaching skills, but also her/his judgments about what to expect from the students. S/he may demand more, or s/he may realize she demanded too much for the allotted time and cuts back or divides the lesson into two sessions. In this way, each lesson, in both the planning and implementing, provides information for the next lesson--affecting both efficacy and outcome expectations.

Science teachers, both preservice and inservice, have been singled out for study numerous times and in varying contexts, one of which is teacher efficacy. Bandura (1997) pointed out that “Teacher efficacy in science education is of particular concern, given the increasing importance of scientific literacy and competency in the technological transformations occurring in society” (p. 242). He exhorted other researchers to link studies of teacher efficacy to knowledge domains, such as science, rather than expect a “one size fits all” instrument to have meaningful predictive power. Bandura suggests that such broadly designed instruments are likely to have underestimated how much teacher self-efficacy beliefs contribute to student achievement.

The studies reviewed here represented several issues in science education. The first issue involved the identification of teacher characteristics and attitudes and their relative importance to student achievement. The next step in the process was to determine the teacher attitudes and beliefs associated with the characteristics that were rated at the highest levels of importance relative to student achievement. Instruments were designed and administered that specifically examined the attitudes and beliefs of science teachers. The next issues dealt with identifying subscales and examining results, and the final issues concerned the interpretation of the results and the implications for science teacher education programs.

In 1975, Lawrenz reported on a study that examined teacher characteristics and attitude and their relationship to student achievement. The purpose was two-fold: (a) to determine the extent of the relationships between teacher characteristics and attitudes and student outcomes/achievement, and (b) to indicate in some way the relative order of importance of the characteristics. Although Bandura (1977, 1997) went to great lengths to differentiate a sense of efficacy from confidence, self-esteem, attitude, and self-concept, the teacher characteristics identified by Lawrenz seem to be logical precursors to both Bandura's work and that of subsequent researchers. One aspect of the rationale for this connection was the inclusion of items pertaining to teacher involvement in professional development activities. The inclusion of such items implied a certain level of professional reflexivity, judgment, and desire to create or influence a particular real or imagined outcome. This recursive process has been identified as a cornerstone of efficacy descriptions and studies.

The focus of Lawrenz's (1975) work was a group of 236 secondary science teachers from 14 states: 84 taught biology, 41 taught physics, and the rest--111-- taught chemistry. Lawrenz primarily used pre-existing instruments to generate data for the study. All 236 teachers completed the following instruments:

1. a questionnaire specifically designed for the study to gather demographic data and other information;
2. The National Teacher Exam (NTE) in Science (from Educational Testing Service) and the Science Process Inventory (SPI), to measure content knowledge, and
3. The Science Attitude Inventory (SAI) to collect data on teacher attitudes because of the documented evidence of connections between teacher attitude and student achievement (1975).

Since Lawrenz (1975) intended to rank the teacher characteristics, criteria needed to be established for what constituted a higher or lower rank. The criteria were the scores of student achievement and attitude, and, to that end, each teacher randomly selected one class of students to complete four instruments: the Learning Environment Inventory (LEI), the Test on Achievement in Science (TAS) compiled from the National Assessment of Science items, and student versions of two instruments the teachers used--the SPI and SAI. This completed the cycle of teacher characteristics, teacher behavior, and student achievement.

To identify teacher characteristics, Lawrenz (1975) looked to the requirements of science teacher preparation programs. She noted that all preservice science teachers were required to take courses in their prospective subject areas, and their knowledge level on the subject area was measured by the National Teacher Examination in Science and the Science Process Inventory. Lawrenz compared the teachers using number of years of teaching experience, number of science methods courses taken, the extent of their content knowledge, their responses about classroom environment, and their strength with regard to the self-improvement variable. The number of courses in teaching methods and the number of years of teaching experience were obtained from the researcher--developed questionnaire, as were questions about activities relating to self-improvement. For example, attempts at professional self-improvement, participation in professional organizations, graduate work, and use of classroom self-evaluation procedures were perceived to be related positively to student learning, and so were included. The teacher's attitudes about science and teaching were obtained by using the scores on the Science Attitude Inventory. The teachers' perspective on the social climate of the classroom was measured by two scales in the Learning Environment Inventory--Formality and Goal Direction. A third scale, Democratic, was determined to be relevant because of the emphasis on student-controlled learning that was prevalent at the time.

When data from both teachers and students were collected and compared, nine teacher characteristics emerged. They were (a) the *NTE-Science* score, (b) the *Science Process Inventory* score, (c) the number of credits in science teaching methods, (d) the number of years of experience, (e) the *Science Attitude Inventory* score, (f) the self-improvement score, (g) the *Formality* score, (h) the *Goal Direction* score, and (i) the *Democratic* score. Lawrenz was interested in the correlations between the identified teacher characteristics and the student scores for the *Test on Achievement in Science*, *Science Process Inventory*, and *Science Attitude Inventory.* The analyses attempted to determine which teacher characteristics showed the strongest predictive value for student achievement and attitude.

As Lawrenz (1975) continued the analyses, patterns of correlations emerged. The teacher characteristics that were most highly correlated (n=235, p< .05) to student science achievement and attitude toward science were the teachers': (a) knowledge of science processes, (b) knowledge of science content information, (c) inclination toward self-improvement, and (d) levels of formality and goal direction.

The canonical correlation between the set of teacher characteristics and the set of student characteristics was .61, and Lawrenz (1975) reports that no others were significant to that level. Lawrenz found that scores on the Science Process Inventory were the best predictors of student achievement across the three student tests of achievement and attitude, but the teachers' SPI scores were not so strong on the canonical analysis or in supplementary regressions completed for individual science courses. The two LEI scales indicated that a class that was goal--directed had high student achievement, but goal directing should not become too formalized into strict rules. If the class becomes too formal with strict rules, the formality scale showed an inverse correlation to student achievement--as formality continues to increase, at some point, student achievement is adversely affected. Another comparison showed attitude as directly related to formality and inversely related to goal direction, but a third analysis provided the opposite result. When results such as these showed ambiguity, Lawrenz indicated the need for clarification of the relationships through further research.

In the results, Lawrenz (1975) identified five variables as the most important: (a) Self-improvement, (b) the score on the National Teacher Exam, (c) the LEI scores on the Formality and (d) Goal Direction subscales, and (e) the Science Process Inventory score. The most consistent predictor of improved student achievement seemed to be the self-improvement variable. The components were all related to the teacher's desire to improve within the teaching profession. Lawrenz describes the emergent profile of the teacher most likely to have students who are achieving at a high level. The single adjective Lawrenz selected was "ambitious" and was described with the following traits: the ambitious teacher (a) ranks high on the professional self-improvement variable, (b) is proactive about improving teaching techniques, (c) is more confident in her/his teaching abilities, and (d) is less likely to require a strict, formalized classroom environment. The ambitious teacher strives toward well-defined goals and is likely to be flexible in order to deal with different situations, both positive and negative, and the classroom environment would likely be less strict and more apt to be changeable. Lawrenz states, "Its [the self-improvement variable] high relation to student outcomes was consistent with other research on teacher characteristics, which generally holds that the teacher's personality is one of the most significant teacher variables relating to student outcomes" (p. 18).

How is Lawrenz's (1975) research connected to studies about science teachers' self-efficacy beliefs that are more recent? An examination of Lawrenz's work through the lens first of

Rotter's (1966) locus of control model and then Bandura's (1977) model of self-efficacy and outcome expectations yields some interesting comparisons. First, a comparison to Rotter: as previously discussed in detail, the locus of control model uses an internal/external continuum representing the degree to which a person ascribes situations and events as being within the scope of her or his control or outside it. In Lawrenz's description of the ambitious teacher, the characteristics can be categorized along the continuum. For example, being proactive about improving teaching techniques, striving toward well-defined goals, and flexibility in dealing with classroom situations can all easily be perceived as within the teacher's locus of control. Characteristics reflected in the scores of the SPI varied, possibly because of differences between instructors' styles of presentation and classroom interactivity in their science methods classes, resulting in science teachers exhibiting differing instructional strategies regarding their own classrooms. This variation could be considered as outside the preservice teacher's locus of control. The comparison cannot be very deep because Lawrenz's work was not designed with Rotter's model in mind, and it would be unfair to retrofit Lawrenz's findings to Rotter's continuum. However, the connections between Lawrenz's work in 1975 and the locus of control model as well as Bandura's (1977, 1997, 2001) efficacy research are quite strong and help to substantiate two points. First, there is considerable research that examines teacher characteristics and their connections to student performance and achievement. Second, teacher characteristics--and beliefs--are related to student achievement, some more so than others are, and some characteristics are in a positive relationship with student achievement, others in a negative one. Studies over the past three decades, such as those completed by Ashton (1982, 1984); Bandura (1977, 1995, 1997, 2001); Bleicher (2004); Dembo and Gibson (1985, 1986); Plourde (2002a, 2002b); Tschannen-Moran and Woolfolk Hoy (2001); and Wenner (1993, 1995, 2001) have all contributed to expanding the understanding of that relationship. The next step is to look at the development of a particular, influential instrument designed to measure characteristics of science teacher efficacy beliefs.

The Science Teaching Efficacy Beliefs Instrument (STEBI)

Riggs and Enochs (1990) elected to continue in the pattern of Gibson and Dembo's 1984 TES and maintain the two dimensions of general teaching efficacy and personal teaching efficacy as distinct variables. Items from the TES were modified to describe situations found in elementary classrooms, to include language that reflects science and science education, and to maintain a balance between positive and negative language orientation. Since several items that loaded on both factors were deleted, additional items were created to enlarge and equilibrate the pool. The resulting 50 items were scrutinized by a panel of judges whose knowledge of the construct was deemed an appropriate criterion for reviewing the items. The scales from the original TES--personal efficacy and teaching efficacy (which Gibson and Dembo stated were equivalent to Bandura's self-efficacy and outcome expectancy, respectively)--were renamed as the Personal Science Teaching Efficacy (PSTE) Scale and the Science Teaching Outcome Expectancy (STOE) scale. The judges classified the items into either of the scales. They were also asked to rate the construct validity of the two scales; items rated differently by three of the five judges were eliminated.

In what Riggs and Enochs (1990) refer to as "a try out study" (p. 628), 71 practicing teachers enrolled in a graduate level course responded to the preliminary draft of the STEBI. As with the TES before it, the STEBI results were mixed. The first factor in each measure--Personal Teaching Efficacy and Personal Science Teaching Efficacy Belief scales, respectively--was clearly supported by the selected items. The second factors in both the TES and the STEBI--the General Teaching Efficacy scale and the Science Teaching Outcome Expectancy (STOE) scale--were much more problematic. Item analyses supported the conclusion that many of the items associated with STOE contained flaws. Eventually, items for the STEBI were retained based on how they loaded on the factors PSTE and STOE, and the eliminated items were examined for patterns. A first pattern of unclear wording emerged when some items cross--loaded on both factors, and a second pattern among the items revealed references to parents or family rather than to teachers as the entities responsible for outcomes. Items matching either of the two patterns were eliminated from the pool, and new items with a negative tone were added to maintain the balance between positive and negative wording orientation.

After the administration of the preliminary instrument to 71 practicing teachers, Riggs and Enochs (1990) found that only 65 responses had complete data sets. When a series of analyses were run on the data from the small sample of teachers, more items were deleted. In the PSTE scale, 22 of 24 items attained a corrected item--total correlation of 0.42 and above, and the bottom six were removed from the pool, resulting in a corrected item-total correlation of 0.50 or more for each of the remaining items. A factor analysis revealed that only three items correlated closely with both scales, and they were subsequently omitted. A final set of analyses of the modified list of items resulted in an alpha of 0.91 and corrected item-total correlations of 0.50 and above for all items.

The STOE scale had lower numbers all around. A reliability analysis showed an alpha of 0.74, and item-total correlations revealed "many weak items" (Riggs & Enochs, 1990, p. 629). The authors included neither a description of the number of items identified nor how many were removed from the pool of items for this dimension, but they state that results of a factor analysis were used to make the selections. A second round of reliability analyses resulted in an alpha of 0.73, and "corrected item-total correlations were raised to 0.37 and above" (Riggs & Enochs, p. 629). Again, without revealing the numbers, the report states, "Factor analysis for the revised scale was much improved, with all items correlating highly with their own scale" (Riggs & Enochs, p. 629).

Riggs and Enochs (1990) used the refined STEBI in a major study involving 331 practicing elementary teachers. The demographics of the group were described as representing both urban and rural schools, both male and female teachers, three racial groups (white, black, and other), grades K-6, a range of years' teaching experience, and a range of district sizes. Indeed these categories were represented; however, on examining the breakdown of the numbers of teachers in each category, it becomes apparent that the vast majority of participating teachers were white (98%), female (88%), and taught in districts with more than 10,000 students (75%). The numbers of teachers in each of the grade levels 1-6 were roughly equivalent as were the numbers of teachers in the subcategories representing the years of teaching experience--except for those with more than 30 years in teaching (3%).

Without stating the number of items in the final pool, Riggs and Enochs (1990) reported several more analyses that resulted in items being removed for not having a high positive discrimination index, having cross loaded or loaded on the wrong factor, or having a corrected item-total correlation of less than 0.53 (PSTE) or less than 0.34 (STOE). The final version of the STEBI scale included 25 items, 13 items on the PSTE subscale and 12 items on the STOE subscale.

Riggs and Enochs (1990) seem satisfied that their original two subscales were supported through the series of analytic procedures, and accept that the lower numbers associated with the STOE scale are consistent with research using Gibson and Dembo's (1984) TES in which the second factor consistently was more weakly supported. Their explanation for the lower reliability suggests that it may be because multiple variables contribute to this aspect of the construct and that at different times a teacher may respond to the same item in different ways. Their rationale also suggests that teachers may perceive some of the situations described by the STOE items as being "external factors over which they have no control" (p. 633). It appears that the authors, who claim allegiance to Bandura's definition and description of efficacy, have used the language of

Rotter in their rationale. For example, the phrases “external factor” and “over which they have no control” are consistent with Rotter’s *Locus of Control* theory and his use of a continuum as a metaphor for the location of a degree of control a person believes s/he has in a given situation. Rotter does posit that a person’s perception of control can vary according to the circumstances or the context of the situation, but his theory does not address the recursive process that Bandura calls reciprocal causation. Although Bandura repeatedly accentuates the differences that separate Rotter’s and his own theories, researchers such as Riggs and Enochs continued to use Gibson and Dembo’s (1984) Teaching Efficacy Scale (TES). Difficulties with the second factors, dealing with more general or global beliefs, are thus perpetuated, and perhaps magnified, each time the original instrument is adapted for a different population.

The Teaching Efficacy Scale (Gibson & Dembo, 1984) has undergone various modifications to suit its application to different populations of teachers. An earlier discussion detailed the development of an instrument that was designed by Gibson and Dembo to measure teacher efficacy in a more global context. The Teacher Efficacy Scale was not the only instrument designed for measuring teacher efficacy. However, numerous modifications of it (more than any other instrument) have appeared in teacher efficacy research over the years. Because of the direct descendancy from the Teacher Efficacy Scale (TES) to the Science Teacher Efficacy Beliefs Instrument (STEBI) developed by Riggs and Enochs (1990), an elucidation of the underlying premise and subsequent findings of the STEBI will be presented.

Along with the changes in the instrument’s design, the name of the instrument underwent metamorphoses as well. For example, the instrument Riggs and Enochs (1990) first used was called the Science Teaching Efficacy Beliefs Scale but had the acronym of STEBI [“I” for Instrument.] The initial version was used with inservice teachers and was referred to as STEBI-A to prevent its being confused with the version used with preservice teachers specifically, the STEBI-B. Other versions of the instrument were developed for measuring teacher efficacy beliefs toward science teaching and learning concerning ethnicity, language minorities, gender, and socioeconomic factors (Ritter, Boone, & Rubba, 2001). Enochs and his colleagues (Enochs, Smith, & Huinker, 2000) modified the STEBI for use with preservice elementary teachers who anticipate teaching mathematics, so it used language consistent with mathematics.

Bleicher (2004) specifically mentions that the "validity and reliability of the instrument was assumed to be intact" (p. 384) when researchers adapted it for use in various studies. Enochs and Riggs established the internal validity and reliability of the instrument in 1990, but no researcher revisited the internal validity and reliability of the STEBI-B until Bleicher undertook the task.

Besides re-examining the factor analysis structure that supported the original pair of scales in the STEBI-B--personal science teaching efficacy and outcome expectancy--Bleicher (2004) was also interested in characteristics and demographics of the preservice teachers that showed strong correlations to either or both of the subscales. He investigated whether any relationships existed between perceived self-efficacy and outcome expectancy and teacher demographics such as gender, ethnicity, age, teaching experience, science courses, and school science learning experiences.

The first administration of the STEBI-B involved 290 preservice teachers enrolled in multiple sections of an elementary science methods class at a large urban university in southeast Florida. After the first administration, revisions were made, and the modified version was then administered to 86 different students with demographic characteristics similar to those of the first group.

The data from the first group of participants were collected before their science methods class began. Bleicher (2004) adhered to the same two subscales, as did Enochs and Riggs (1990): (a) the PSTE with 13 items measuring self-efficacy and (b) the STOE with 10 items measuring outcome expectancy. Items 10 and 13 were set aside, not used in this phase of the study, and were later modified. The total PSTE and STOE scores were used in the analyses of possible associations with the demographic variables listed previously. Bleicher used the Statistical Package for the Social Studies (SPSS) for the factor analysis. Both oblique minimum and varimax (orthogonal) rotations were used to compare factor loadings and correlations. Factor loadings were similar with both rotations, and in order to compare the loadings with the Enochs and Riggs information, it was decided to use the oblique rotation with Kaiser Normalization. To examine associations with PSTE and STOE scores, the background variables were used, and the data set was separated into the categories composing each demographic variable. Bleicher reported using either an independent *t*-test or ANOVA to detect any significant differences between group means. The independent *t*-test was used with variables with only two groups and ANOVA was used for variables with more than two groups. For example, Bleicher reports, "… for gender the PSTE and STOE scores for the two groups, male and female, were compared using a two-tailed, independent *t*-test" (p. 386).

In the comparison of results from his study and the Enochs and Riggs (1990) study, Bleicher (2004) found general homogeneity of the two scales between the studies. Bleicher states, "The means and standard deviations of all 23 items agreed remarkably well with the original data collected by Enochs and Riggs (1990)" (p. 386). The Bleicher study showed similar loadings as the original analysis of the 23 items. "The two factors accounted for 36.38% of the variance, with eigenvalues of 5.30 and 3.14 for PSTE and STOE, respectively. Bleicher reported that the two factors (PSTE and STOE) were moderately correlated ($r = 0.12$ at a delta value of 0), signifying that "they are related but independent constructs" (p. 386). The following rationale for deleting Item 10 and Item 13 followed.

Two items, numbers 10 and 13, produced questionable values, according to Bleicher, the item loadings and the results were somewhat inconsistent between the two studies--Bleicher's (2004) and the one from Enochs and Riggs (1990). For example, Item 10 exhibited crossloading in the Enochs and Riggs study but not in the Bleicher study. The loading of Item 10 on STOE of 0.31 on the Bleicher study was slightly below the cutoff level of 0.32, which he used. In both studies, (Bleicher's and Enochs and Riggs's) Item 13 demonstrated both cross-loading and low values. In 1990, the PSTE of 0.14 and STOE of 0.22 and in 2004, the PSTE of 0.10 and STOE of 0.13 put the Item 13 values well below the recommended 0.32 minimum to which Bleicher adhered. The two items, 10 and 13, contained a difference in wording from all other items.

Bleicher cites the following specifics as examples:

Item 10: The low science achievement of **some** students cannot generally be blamed on their teachers.

Item 13: Increased effort in science teaching produces little change in **some** students' science achievement (2004, p. 387).

The inclusion of the word "some" in both items was used as a qualifier for "student." Bleicher (2004) surmised that the inclusion of the qualifier words affected how respondents reacted to each of the statements. When students were interviewed following the administration of the instrument and asked about their interpretations of Items 10 and 13, this perspective was substantiated. The two items were modified by omitting the word "some" in each item, and the revised STEBI-B was administrated to a new set of 86 preservice teachers on the first day of their elementary science methods class.

The data collected in the second administration of the instrument were analyzed using the same statistical techniques as the first set of analyses. For the PSTE subscale, the results showed no significant difference between the factor analyses of the first and second data sets of the study. All items besides 10 and 13 on the STOE subscale exhibited similar factor analyses and reliability statistics when compared to the first data set in Bleicher's (2004) study. Item 10 exhibited a factor loading of 0.64 on the STOE and 0.30 on the PSTE, which Bleicher claimed indicated a clear loading on the STOE, and the item-total correlation for Item 10 of 0.53, was well above the 0.25

value in the first study. The factor loading of Item 13 was 0.68 on STOE and 0.24 on the PSTE in the second study, which Bleicher claimed indicated, as with Item 10, a clear loading on the STOE. The item-total correlation for Item 13 in the second administration of the instrument was 0.47, which, again like Item 10, was an improvement over the item-total correlation of Item 13 in the first administration of the STEBI-B.

Bleicher (2004) concluded that the basic integrity of the original PSTE and STOE scales was intact, but that two items numbers 10 and 13, displayed problematic results in both the factor and reliability analyses. Teachers' comments regarding Items 10 and 13 referred to the use of the qualifying term "some" preceding "students" and the ambiguity it created. When the word "some" was removed from the item language and the modified items re-inserted in to the instrument, the intention of the item was clarified and the reliability of the instrument increased. Bleicher stated that additional research to reinforce the idea and level of increased reliability was underway at the time of publication. Bleicher also cautioned other researchers who use the STEBI--B to consider conducting a factor analysis to ensure that items load on the two factors of PSTE and STOE as expected.

The work of Rotter (1954) and Bandura (1971, 1997) serves as the theoretical underpinning for the development and refinement of general teacher efficacy scales such as Teacher Locus of Control Scale, Teacher Efficacy Scale, and Ohio State Teacher Efficacy Scale as well as subject-specific scales like the Science Teacher Efficacy Beliefs Instrument. Tracing the evolution of the construct of teacher efficacy and the development of instruments to measure it is foundational to the development of a standards-based science teacher efficacy instrument, the goal of the current study.

CHAPTER 3

METHODOLOGY

Introduction

The sections of this chapter present the stages in the evolution of the Standards-Based Science Teacher Efficacy Instrument and its analysis related to validation, predictive power, and discriminating sensitivity. The instrument development and the analysis of the data collected occurred in three stages. The first stage grew from an informal discussion in a science methods course, and the initial instrument was developed and piloted, in the second stage. The information collected from preservice and inservice teachers who participated in both the first and second stages exerted an immense influence on the content and design of the final instrument that was used in the third stage.

Each set of participants is described from information they provided, with more details available as the study progressed. The analytic methodology is described and rational, logical arguments are presented for decisions made at critical points in the instrument's development, the analyses of the data, and conclusions drawn from the information collected. Both descriptive and inferential statistical methodologies were performed and are briefly described in a data analysis section at the end of this chapter.

Initial Investigation

Subjects

A group of 12 preservice teachers in a science methods class at a regional campus of a major Midwestern university expressed concerns about teaching science. They expected to be teaching science at the first through eighth grade levels within a prescribed elementary or middle childhood level curriculum; most would be in situations where science was only one of their several teaching responsibilities; few aimed for careers in departmentalized curriculum settings; and fewer still would focus totally on science. Few of the preservice teachers thought they had adequate preparation in science content studies to teach science. The majority had few positive experiences as students in high school science classes and previous university science courses; nonetheless, they expressed a desire to teach science effectively and wanted to know how to overcome feelings of being ill prepared. A decision was made by members of the group to explore their desire to be effective science teachers by agreeing to dialogues with the instructor as a means of guided reflection. A short series of interview questions emerged from an expeditious review of self-efficacy research, especially the work of Enochs and Riggs (1990) and their development of the Science Teaching Efficacy Beliefs Instrument (STEBI).

The questions were devised as a guide to elicit reflective comments, not as a strict template for the collection of qualitative data; therefore, the direction of the questioning depended in large part on the responses of the preservice teachers. For example, if a response to the question "How do you feel about science?" included elaboration of a positive behavior of a teacher, it was used as a springboard for further discussion at that time, rather than waiting until its appearance in the sequence. The following list includes the questions that were common to all participants of the discussion. Other queries were spontaneous and not recorded.

1. What do you think about science, intellectually? Explain.
2. How do you feel about science? Explain.
3. Do you believe the negativity or positivity of the experience was related to the science topic itself or to something else? Explain.
4. How do you feel about teaching science? Explain.
5. What do you think a teacher does that makes science class experiences positive for students?

6. What do you think a teacher does that makes science class experiences negative for students?
7. What do you believe about your ability to teach science effectively?

Initial Investigation Analysis

The emphasis of the exercise in reflexivity focused on the participants' beliefs and feelings about science and science teaching. They did not focus their attention on performance expectations of students. It was more relevant for the prospective teachers to examine their personal perspectives on teaching science before examining standards for science classrooms and curricula, and the teaching expectations related to standards. The participants quickly recognized that one of the main reasons for their uneasiness about teaching science revolved around the level of intimidation they felt about teaching such a complex subject as science. A second reason was their level of knowledge and understanding of science concepts--both basic and complex.

A third reason also emerged that was related to both the intimidation level and to the levels of science content and pedagogical knowledge they believed they had. The prospective teachers were unsure of what constituted effective science teaching. Because their own science learning experiences were dominated by passive reception of an instructor's lecture notes and lack of interactive experiences, they were uncertain about what was involved in teaching science in an environment of standards-based instruction and learning that exemplified a more progressive approach to science education. A motivation emerged to attempt two goals: (a) to identify what constituted standards-based science education, and (b) to identify how to help preservice science teachers feel less intimidated by science and the idea of teaching science. The parameters of an efficacy instrument began to take shape.

Suggestions made by the preservice science teachers were to examine more closely how participants felt or what they believed about science as well as what they felt and believed about teaching science. One preservice teacher's insight was that "… how the participants felt about science [i.e., their emotional response] affected how well they thought they would be able to teach it" (Kerr, 2002, p. 2). The student reported that she was somewhat "afraid of science" and did not think she would be able to teach it well. After several probing questions by the instructor, she remembered that she and her father had spent a great deal of outdoor time looking at and talking about rocks and rock formations. What she had originally categorized as fun time with her dad actually influenced her interest in geology and prompted her to take several science courses. She had not connected the experiences with her father, an amateur rock hound, with either her taking geology classes or her ability to create an interesting science lesson. When she made the first tenuous connections, she immediately realized that she knew more than she thought she did, and that teaching science might not be so intimidating after all.

A second set of suggestions involved item language and criteria. For example, the preservice teachers cautioned that items not be too narrow, ambiguous, or irrelevant. Future respondents needed to be sure of what exactly was being asked of them. This set of suggestions led to the most significant revisions of the short questionnaire.

The first challenges were to identify research relevant to science teacher self-efficacy beliefs and to review the findings. A decision was made to find the self-efficacy instrument most often used or cited in efficacy research about science teachers or science teaching. The reasoning was that the instrument most often used would likely have considerable data to support its validity and reliability. The Science Teaching Efficacy Beliefs Instrument (STEBI), developed by Enochs and Riggs (1990), was the basis for the set of interview questions used with the preservice teachers, because it provided the designer with a frame of reference for the essence of the questions. It turned out to be the most durable of the various instruments over time and was widely used and modified among researchers.

When information about STEBI was first published, its statistics were published in the journal, *Science Education,* and claims were made that the instrument was both reliable and valid. Two scales that emerged were named Personal Science Teaching Efficacy (PSTE) and Science

Teaching Outcome Expectancy (STOE). Cronbach's alpha for the PSTE scale was reported as 0.92, and all items showed corrected item-total correlations of 0.53 or higher. Cronbach's alpha for the STOE scale equaled 0.77, with corrected item-total correlations of 0.34 and higher. Eigen values of 6.26 accounting for 25.0% of the variance, and 2.71 accounting for 10.8% of the variance, were reported for PSTE and STOE, respectively. Construct validity claimed to be enhanced by the homogeneity within and distinctiveness between the scales. This was supported by the report of high correlations of the items among themselves within each scale and lower correlations between the two dimensions.

Bleicher (2004) conducted a study to re-examine the reliability and validity of the Enochs and Riggs instrument of 1990. Results were similar and supported the basic integrity of the instrument. The single recommendation Bleicher made was to change the wording for two items by removing the word "some" as a qualifier before the word "students." Although the alteration slightly improved the reliability of the items, no other changes were made or suggested to reflect the shift towards standards--based science education for both K-12 students and science teachers.

Pilot Study of Initial Instrument

Pilot Study Subjects

After perusing relevant literature, several characteristics emerged regarding the participants in various studies. First, the number of studies involving elementary teachers as subjects of self-efficacy research was far greater than middle or secondary teachers. Second, secondary science teachers have a considerably stronger and more thorough academic preparation in a science discipline than middle level teachers or elementary teachers. The confidence they feel in their level of knowledge and understanding of science concepts often affects their sense of self-efficacy in a positive direction (Simmons et al. 1999). Third, the researcher's personal experience with science curriculum, assessment, and instruction at all grades--K through 12--indicated that students' experiences with science in the middle grades were often dichotomized, and they eventually cemented their attitudes toward science in subsequent years (American Association of University Women, 1992). This idea led to questions about teacher use of strategies and activities in middle level science. This interest in how students' innate curiosity about the natural world was either enhanced or diminished during the middle grades led to an examination of middle level teachers' attitudes and beliefs related to science. Research literature reported on teachers' sense of efficacy related to teaching in general; however, the variety of circumstances among science teachers and within individual teachers' situations affected data and findings related to teachers' sense of efficacy about their own teaching (Posnanski, 2002; Roberts et al., 2001).

As in the preliminary investigation, preservice teachers would be the respondents in the pilot study, but the target group for subsequent research was more clearly defined. The group was made up of graduate students in an Adolescent and Young Adult (AYA) teacher preparation program at a large Midwestern university; all were in a fifth-year Master's degree program. Two cohorts of approximately 30 each were invited to participate. Unlike the respondents in the initial investigation, the preservice teachers were not in an elementary program. All were working toward completing requirements for licensure at the secondary level, which included grades 7-12. The instrument was administered during a regularly scheduled science methods class and required approximately 20 minutes to complete.

Pilot Study Instrument Development

The survey used with the secondary level preservice teachers consisted of the following: a page asking for demographic information and three pages containing 44 items related to beliefs, and an open-ended request for feedback on item clarity. The requested demographic information related to licensure area, number of science content courses, number of science methods courses, and whether the respondent's status at the university was on a full-- or part--time basis. The items were based on teacher behaviors supported by national science education standards and/or science teacher preparation program standards. The open-ended item provided an opportunity for item review by the preservice science teachers.

The first step in the development process was identifying components or issues that were fundamental to the standards. Since significant changes in curriculum, instruction, and assessment had occurred since the instrument published by Enochs and Riggs (1990), it was deemed prudent to generate as many new items as needed rather than simply replicate the STEBI with only minor modifications. Some items were similar to those in the STEBI; others were completely different. The items in the pilot instrument mirrored language consistent with documents such as the *National Science Education Standards* (NRC, 1996) published by the National Research Council; and *Project 2061: Science for All Americans* (AAAS, 1989), and *Benchmarks for Science Literacy* (AAAS, 1993) published by the American Association for the Advancement of Science. The language also was consistent with standards-based curriculum frameworks (Council of Chief State School Officers, 2002) and other guides to standards-based science education published by professional organizations, such as *Pathways to the Standards: Elementary School Edition* (Lowrey, 2000) published by the National Science Teachers Association. During the development process, items were written, scrutinized, and revised to maintain equilibrium among the aspects of positive and negative language orientation and between items relating to teacher beliefs about self-efficacy of science teaching and beliefs about their relationship to student achievement. The items in Table 3.1 are examples of new items developed for the pilot survey.

New Items Developed for Pilot Survey	
Teaching Self-Efficacy Beliefs	Beliefs About Student Achievement
• I am not very effective in developing science experiments for students. • I am prepared to plan for inquiry-based science lessons. • I feel capable of teaching science to all students. • Issues of safety in the science classroom do not intimidate me.	• Even if I increased my effort in teaching science, some students will show little change in achievement. • A student's understanding of science concepts is best measured by how well s/he performs on objective tests. • Using technology in science class ensures that students will have higher achievement. • Students will not be successful in science if they work in groups.

Table 3.1: Pilot sample items for teacher self-efficacy beliefs and beliefs about student achievement.

After developing 44 items based upon science education standards and reviews by experienced science educators, a Likert--type response scale was included. The 5-point scale included Strongly Agree (SA), Agree (A), Neutral (N), Disagree (D), and Strongly Disagree (SD). The highest value--5--was assigned to the most desirable response and the lowest value-1-was assigned to the least desirable response. The full pilot survey is included in Appendix A.

Pilot Study Analysis

Of the 60 surveys distributed to the preservice teachers, only 23 had all items completed. With so few returns, a decision was made to use descriptive analyses and study the content of the responses on all the surveys for common themes. Rather than analyze the instrument as a whole, items relating to the same or similar issue were compared to each other. Although only 23 surveys had all items completed, counting responses for pairs or sets of similar items produced as many as 35 marked responses. For example, the surveys had 12 items relating teacher responsibility to student achievement--six items paired positive student achievement to teacher responsibility, and six items paired negative progress to teacher responsibility. Thirty-five surveys had all 12 items marked; this allowed for comparisons between similar items.

In each case, the items were examined individually to determine the direction of the strongest level of agreement that was most closely aligned with standards-based philosophy. For example, in some items, the most closely aligned level was “Strongly Agree,” while for others it was “Strongly Disagree.” A tally was made of the number of responses for the two strongest levels of agreement or disagreement for each item in the pair or set, and the tally was compared to the total number of responses available for that pair or set of items. The resulting percentages provided some insight into what the participants believed.

Pilot Study Results

The following paragraphs continue a thread previously mentioned about items relating to teacher beliefs regarding student achievement and describe the comparisons. In general, the respondents considered the teacher responsible for positive student achievement, but not responsible for students’ lack of progress. For the 12 items under scrutiny, 35 preservice teachers responded to all 12, and the combined maximum possible score of each set of 6 items was 210 points. Of the possible 210, 159 points or 75.7% of the responses were either the highest or the second-highest value in support of teachers being responsible for positive student achievement. Totals for the 6 items related to teacher responsibility for negative student achievement showed that 130 of 210 points, 61.9% of the responses, were at the two strongest levels indicating the belief that teachers were not responsible for lack of student progress. While the numbers might seem high, the trend is supported by other efficacy studies (Dembo & Gibson, 1985; Enochs, Riggs, & Scharmann, 1995; Goddard, Hoy, & Woolfolk Hoy, 2000; Muijs & Reynolds, 2002).

A second example involves the following two items:

1. Student achievement in science will improve when students work in groups.
2. I feel at ease when students are doing group work in science experiments and investigations.

After tallying the responses for these two items, it was found that 42 returned surveys had both items marked. When the responses for the two highest levels of agreement on each item were counted, it was determined that 32 of 42 (76.1%) believed that achievement would improve if students in science class worked in groups. Results for the other item, however, showed that only 24 of the 42 responses (57.1%) indicated a level of comfort in having students work in groups. These two items had logical connections--a corresponding item that could easily be used in a comparison.

Other combinations of items were less clear-cut and did not lend themselves to further analysis. For example, safety in the science classroom, inviting the principal to the science classroom, and using technology addressed three issues separately and uniquely. Each had no corresponding item for comparison. Consequently, the responses to the open-ended request for feedback on the survey were studied for comments about modifying the item language or content.

Some participants added comments or notes in the margins of the survey near specific items, and some made notes at the end of the survey, but most did neither. A second detraction was the dearth of complete surveys. Too few participants responded to all items, and this, in turn, prohibited a valid statistical analysis because of missing data. An unanticipated opportunity arose, however, to ask questions of several participants, and the comments provided insight into the teachers' thinking as they responded, and the information enabled revisions to several aspects of the instrument.

Participants' responses revealed that several thought some items were too vague. They also reported that the idea of an investigation into teacher self-efficacy beliefs based on standards was a good one, since they anticipated being expected to create a learning environment and teach in a standards-based classroom.

Pilot Study Conclusions

After some participants reported that the items were vague, each item was scrutinized from several perspectives. The first was to ascertain that the item did not have a compound structure. Each item should address a single topic, or it would not be possible to assign a single score to the item. Another criterion involved maintaining a balance between items with positive wording and those with negative wording. A third filter also involved maintaining a balance-between items relating to teacher self-efficacy beliefs and those relating to beliefs about teacher responsibility for student achievement.

An additional conundrum arose after studying the comments made by the preservice teachers. Several found it difficult to express beliefs about behaviors they anticipated, rather than about situations with which they had experience. They also found a conflict between beliefs referring to teaching in general and beliefs about their personal behavior as teachers. One of the participants also observed that items were not of equal value or "grain size." Some items referred to somewhat trivial behaviors, while others referred to broader ideas or issues.

Recommendations Emerging From the Pilot Study

1. Data might be more useful if a subsequent study involved practicing teachers rather than preservice teachers. The overall goal of the study was to gather information that might have an impact on programs for preservice science teachers; however, practicing teachers are more likely to have a sense of what is useful in the classroom, and they have practical experiences as their frame of reference when responding to items.
2. Since the purpose of the survey was to examine self-efficacy beliefs of science teachers as well as their beliefs about student achievement, keep all the items oriented to the personal aspect of teaching and not oriented to beliefs about issues or behaviors related to the teaching profession in general.
3. Participants might provide more insightful responses if the items related to their beliefs about specific tasks germane to standards-based K-12 science.
4. Write items that address a single issue or behavior.

5. As much as possible, keep the items equal or equivalent in value.
6. Maintain a balance between items with positive wording and those with negative wording.
7. Maintain a balance between items referencing teacher beliefs about self-efficacy and items referencing beliefs about student achievement.
8. Standards-Based Science Teacher Efficacy Instrument

The first three recommendations generated from the pilot study results: (a) involving practicing science teachers rather than preservice; (b) focusing on personal teacher beliefs rather than beliefs about the teaching profession in general; and (c) linking item development to specific standards-based tasks for the science classroom were fundamental to the conceptual framework and development of the standards-based science teacher efficacy instrument. The last four recommendations were oriented toward item development and maintaining a psychometric equilibrium when assembling the final instrument; they were part of the recursive process of review-and-revise. Attention to those four item recommendations is expanded in Chapters 3 and 4.

Standards-Based Science Teacher Efficacy Instrument

SBSTEI: Issues and Considerations

The first recommendation, involving practicing science teachers rather than preservice teachers, was considered and implemented. Conversations with progressive science teachers influenced the shaping of parameters for recruiting participants. Using experienced teachers had several advantages. One is that they simply had more experiences than preservice teachers. They also had a greater variety of experiences--with students, education bureaucracies, teaching assignments, and standards. During a review phase of an early version of the pilot instrument, it was advised that a change be made regarding the grade levels taught by the participants. The initial investigation was with preservice teachers anticipating certification in grades 1-8. A review of relevant literature revealed a perception that more research was needed with middle and secondary level teachers; therefore, changing from elementary to middle level preservice science teachers was recommended and put in place for the pilot study. Encouraged by remarks from a panel of experts--four science education professors from the university--regarding the responses of the middle level teachers in the pilot study; they were recruited for the new instrument.

Agreement about involving middle level science teachers also required that the term "middle level" be defined for this study. The specifications and parameters concerning the teachers' classroom assignments were the following: (a) science teachers who taught classes with students in grades 5 through 10, and (b) those whose science teaching assignment was at least 50% of their time. This group was recruited through district-level personnel, school-level administrators, and personal contacts via postal and electronic mail.

Within the framework for this study, a triad existed among three issues: (a) standards for teaching, (b) standards for student learning, and (c) teacher beliefs about both teaching and learning. Repeated studies showed a strong consistency in results and interpretations related to teacher beliefs about the profession of teaching. General teaching efficacy was first described in the Rand Study (Armor et al., 1976), expanded by Gibson and Dembo (1984), and supported by Soodak and Podell (1993) and Deemer and Minke (1999). Because of this strength and the desire to focus on specific tasks related to standards, participants would respond to specific behaviors related to student learning. It was logical to generate items targeting beliefs about their own efficacy in teaching toward successful student performance in science. All items related to this parameter were oriented toward teacher efficacy beliefs on a personal level rather than toward teaching in general, and the new instrument would be a collection of items related to teacher beliefs about their own teaching efficacy and their beliefs about student learning relative to their sense of efficacy, and not to teaching in general.

The connection between teacher self-efficacy beliefs and student achievement is a well-supported theme throughout efficacy literature. Generally, the more positive the teacher's beliefs about her/his own ability to teach the subject under investigation, and the more strongly those beliefs are held, the more likely her/his students are to achieve at higher levels. The reverse is also true--students are less likely to achieve at higher levels if their teachers have a less positive sense of self-efficacy about teaching or if their beliefs are held more weakly (Tosun, 2000; see also Ashton, 1984; Bandura, 1971, 1977, 1997, 2001; Gibson & Dembo, 1984; Goddard et al., 2000; Riggs &, Enochs 1990; Roehrig & Kruse, 2005; Woolfolk & Hoy, 1990).

A deliberate decision at this juncture concerned the scope of the proposed survey. A second aspect of a standards-based instrument addressed the idea of student achievement as a consequence of effective teaching. In an attempt to maintain the cohesiveness of teacher beliefs when responding to the survey items, rather than examine student achievement as a separate variable, items were designed to address teacher beliefs about the relationship that might or might not exist between teacher behavior (standards-based) and student achievement.

During the item development process, several issues surfaced about each teacher's professional development--in both the teacher education program and after initial certification or licensure. The issues emanated from both previous research and natural curiosity. For example, Ashton (1984) studied whether information about teacher self-efficacy beliefs could be used to design effective teacher education courses or programs. Other studies and/or instruments looked for the level of influence field experiences or student teaching had on preservice teacher self-efficacy beliefs (Cannon, 1996; Cantrell, 2003; Enochs & Riggs, 1995; Plourde 2002b; Soodak, 1997). Another question was related to teachers' beliefs about student achievement. The standards-based survey contained items that could potentially address those issues.

The survey was divided into four sections. The first section requested the teacher's information about her/his school. The second section collected information about the respondent's teacher preparation programs and experiences before initial certification or licensure. The third section contained items about the teacher's professional development experiences after initial certification or licensure, including participating or presenting at science and science education conferences. The fourth section contained the items relating to teacher beliefs.

In the first section, the demographics page of the survey asked the participant about characteristics of her/his school. The items dealt with each teacher's school and included items about whether their school was

1. Public, private, or other.
2. Urban, suburban, or rural.
3. Large (> 1000 students), medium (400-1000), or small (< 400).
4. A high (>60%), medium (26% to 60%), or low (<26%) percentage of families receiving financial assistance.

The second section consisted of three subsections and requested information about the participants and information about their preservice experiences. The items in the first subsection addressed personal and professional information--the participant's name, the gender of the participant, the most advanced degree completed, the year the most advanced degree was completed, the number of years teaching, and the number of years teaching science at grades 5-10. Items in the second subsection dealt with preservice preparation and included major area(s) of concentration, grade level of initial certification or licensure, type of teacher education program-undergraduate or graduate level, the number of graduate-level hours or credits completed after the most advanced degree, and the types of science classes taught during their careers. No differentiation was made regarding the types of graduate-level courses or classes in which teachers enrolled. Items in the third subsection related to types of inservice professional development experiences after initial certification or licensure. They included the following categories: (a) graduate level classes or courses in a formal setting and (b) attendance or presentation at various levels of science or science education conferences--local/district/state, national, or international.

During an informal, spontaneous discussion after a professional development session unrelated to this study, one experienced teacher in the group mentioned that she made a major change in her teaching strategies after attending a national conference for science teachers. Her motivation and commitment to teaching were renewed. At the time, she was in a graduate degree program for science education, but had not ever attended a teacher conference. This single story prompted questions among the other members of the informal group about who had attended or participated in similar conferences and when. From that discussion, the last two demographic questions emerged. The items asked whether respondents had attended or presented at either a science education or a science conference at any level (e.g., local, state, national, or international) before their initial certification or licensure. On that same page were also statements clarifying the difference between "science" and "science education," as well as the term "conference," as they were used in the survey. The term "science" referred to discipline-specific content such as anatomy, chemistry, or geology, and the term "science education" referred to pedagogy and discipline-specific content information, such as science methods courses. The term "conference" referred to meetings of professionals within a discipline. Science conferences are those designed especially for scientists or content specialists, such as meetings sponsored by the American Chemical Society. Science education conferences referred to those gatherings sponsored by local districts providing information for science teachers about science teaching or nation-wide gatherings such as those of the National Association of Biology Teachers. The thought was that data from these two items might yield interesting results when correlated with other variables and provide an impetus for further research.

SBSTEI: Development Sequence

In the initial investigation analysis section using the Standards-Based Teacher Efficacy Instrument (Enochs & Riggs, 1990), interviews with the preservice teachers revealed that they were unsure of just what they were supposed to do in a standards-based science classroom, how it was different from a traditional science class, and whether they would be effective science teachers. This, in turn, led to further questioning with regard to three ideas: (a) standards-based performance expectations for students; (b) types of activities and strategies teachers use to enable students to meet standards; and (c) teacher self-efficacy beliefs about their own performance of the activities and strategies that will enable students to meet standards. Measuring teacher self-efficacy beliefs would entail several major challenges. The first was naming and describing science classroom activities and strategies, and the second was distilling the list of activities and strategies into tasks. The third was to use the tasks to create a set of standards-based attributes for both K-12 science and science teacher education. The fourth challenge was creating items that sufficiently addressed the attributes for curriculum, instruction, and assessment related to pedagogical principles for a standards-based learning environment and included the incorporation of language that would help respondents differentiate between self-efficacy beliefs and actual behavior. The fifth responsibility lay in ascertaining that the major areas of teaching and learning were thoroughly addressed. The sixth challenge was to construct a viable instrument that would be both valid and reliable. See Figure 3.1 for the development progression for the standards-based science teacher efficacy instrument.

Developmental Progression for Standards-Based Teacher Efficacy Instrument

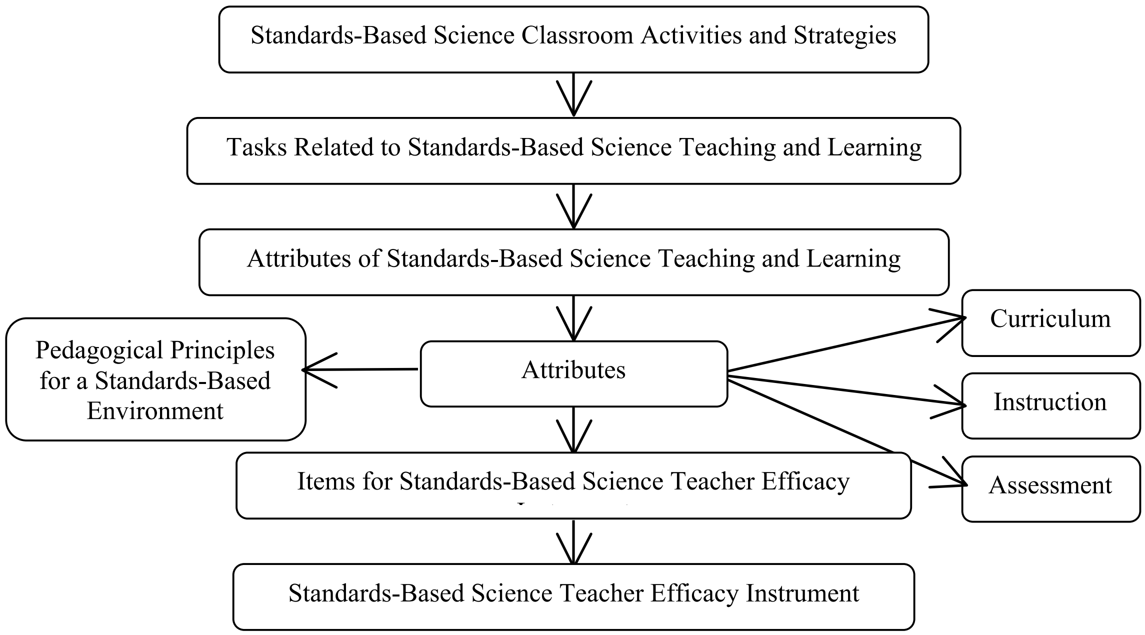


Figure 3.1: Developmental progression for the standards-based teacher efficacy instrument.

From Science Classroom Activities and Strategies to Tasks

Both the *National Science Education Standards* (NRC, 1996) and the *Standards for Science Teacher Preparation* (NSTA, 2004), designed in a collaboration between the National Science Teachers Association and the National Council for Accreditation of Teacher Education (NACTE) describe learning and teaching expectations. The *National Science Education Standards* (NRC) describe standards for K-12, and *Standards for Science Teacher Preparation* (NSTA) includes rubrics for evaluating teacher performance related to the standards. The emphases on performance standards connect logically to the idea of certain activities and strategies being more closely aligned with standards. The identification of activities and strategies characteristic of standards-based science instruction and learning was the initial objective. As such, documents created specifically to further the goals of universal science literacy were analyzed. They were documents such as *Science for All Americans: Project 2061* (AAAS, 1989), *Benchmarks for Science Literacy* (AAAS, 1993), and the *National Science Education Standards* (NRC). Each carried a high level of credibility among science educators. The designers and writers approached the ideas of science literacy and science competency from different cognitive and discipline perspectives but shared major philosophical frameworks, such as an advocacy for more and better interactive science lessons for students at all grade levels.

The connection between activities and strategies and teacher beliefs led to further study of Bandura's (1977) research and development of the concept of self-efficacy. His research emphasized the importance of identifying tasks about which respondents could express their sense of efficacy, and he advocated for as much specificity as possible given the topic and research goals. This gave additional support to the idea of identifying tasks for the proposed instrument.

The strategy used to create the list of tasks involved brainstorming and review sessions with science educators. The group was made up of eight science educators--six of whom were currently teaching, one district-level science specialist, and one graduate student working on an advanced degree in science education. The group was participating in the development of science curriculum materials for a state-funded project and met regularly over an 11-month period. They represented grades 6-12 and all academic science areas in grades K-12 except for Advanced Placement Biology, Chemistry, and Physics; and their schools were in urban, suburban, and rural districts across the state. The brainstorming and review sessions occurred over the life of the funded project, and the first three sessions yielded a list of approximately 60 tasks deemed important to the implementation of a K-12 standards-based science education.

Using a process similar to the one used to develop the tasks related to K-12 science, the group of eight generated a list of tasks related to learning how to teach science based on *Standards for Science Teacher Education Programs* (NSTA, 2003) in order to address tasks that may have been overlooked in the K-12 analysis. Many of the tasks that emerged in the examination of teaching standards were the same as the tasks for K-12 science education. Some tasks, however, such as organizing an interactive science lesson with appropriate materials, keeping students safe in a science classroom during interactive lessons, and designing lessons in a constructivistic frame of reference were exclusive to learning to teach science as described in science teacher preparation documents.

From Tasks to Standards-Based Attributes of Science Teaching and Learning

Focus sessions with teacher educators distilled the lists of tasks-60 from K-12 and 20 additional tasks from teacher education--to a list of 50 standards-based attributes. Each of the attributes makes an essential contribution to the creation and maintenance of a standards-based learning environment for science. Each represents one face of the whole environment; they are not mutually exclusive, and they must work in concert with each other, but they are not all equal in their impact on science learning. The group of six science teachers from the curriculum project, four teacher educators on the faculty of a large Midwestern university and two science curriculum specialists from districts at opposite ends of the state, refined the list of attributes to a set of 22, which were labeled *Standards-Based Attributes of Science Teaching and Learning*. The group created the list with the understanding that it is an on-going recursive process open to revision and refinement.

An evolutionary process that began with preservice teachers' expressions of self-doubt about teaching science culminated in the identification of standards-based attributes of teaching and learning science. Identification of the attributes facilitated the item design for the teacher efficacy beliefs instrument based on science standards. A categorization of the attributes was designed to affirm that no gaps remained in the schema of science teaching and learning. The four categories and their descriptions are as follows:

1. Pedagogy and classroom environment represented the knowledge, skills, and dispositions that are the focus of performance criteria developed in *Standards for Science Teacher Education Programs* (NSTA, 2003), including both science content and pedagogical knowledge.
2. Curriculum represented the attributes most closely associated with the knowledge and selection of science content and processes to be taught and/or learned.
3. Instruction represented the attributes most closely associated with the planning, organization, and delivery of strategies and activities related to science content and processes.
4. Assessment represented the attributes most closely associated with the methods and means of authentication and evaluation of the acquisition, retention, and transfer of science content and processes.

Sorting attributes into the categories indicated how the attributes tended to cluster into broader academic areas; however, a valid, logical argument can be made for certain attributes being reassigned to other categories. Generally, an attribute easily slipped into the designated category, but when there was a possibility of its being included in multiple categories, it was placed where it had the strongest support from the criteria in the definition of the category. For example, the attribute "Lesson planning for inquiry-based science" could be included with "Pedagogy and Classroom Environment." Not only is the development of lesson plans integral to a science Curriculum, but learning how to design standards-based, interactive lessons is also key to the overall assimilation of pedagogical knowledge and skills. It was placed in the Curriculum category because lesson plans present the substance and methods of the enacted curriculum.

At the end of Chapter 1, a discussion was presented differentiating between the attribute categories of Pedagogy and Instruction. In the current study, "Instruction" is operationally defined as the enacted curriculum, the performances that are played out in the classroom-between the teacher and students, between the students and the information, and between the students. Pedagogy refers to the acquired infrastructure of knowledge, understandings, and dispositions that enable teaching and learning to occur to the greatest benefit of both the teacher and student. See Table 3.1 for the complete list of attributes.

Standards-Based Attribute Categories of Science Teaching and Learning
Pedagogy and Classroom Environment • Classroom management for interactive science • Cooperative learning groups: Appropriate use in science • Diverse learners and special populations in the science-learning environment • Inquiry-based science: Defining and describing it • Motivation techniques for reluctant learners of science • Parental involvement: Extending science beyond the classroom • Prior knowledge and experiences related to science: Identifying what students already know • Science education textbooks: Evaluation and selection • Teacher accountability and ethics • Teacher self-evaluation and reflection: Insights and revelations
Curriculum • Concepts in science disciplines: Identifying and teaching concepts appropriate for students • Lesson planning for inquiry-based science • Safety in the science classroom, lab, and in the field • Science experiments and investigations: Designing and/or identifying • Scientific processes: Their role and importance in science • Scientific terminology: Its role and importance in science
Instruction • Hands-on activities: Designing and/or modifying • Science equipment and materials: Selection, and maintenance • Science teaching strategies: Identifying strategies and developing teaching skills • Technology: Its role and importance in science
Assessment • Alternative/Authentic assessments (e.g., creating graphics, writing extended open-responses, or designing projects): Their role and importance in science • Traditional tests (e.g., multiple choice, true/false, and matching): Their role and importance in science

Table 3.2: Standards-based attribute categories of science teaching and learning.

From Standards-Based Attributes to a Standards-Based Science Teacher Efficacy Instrument

Although the list of 22 standards-based attributes emerged from the identified tasks, the possibility existed that some might need more than one item to express different perspectives. For example, one attribute relates to the use of technology in science classrooms. One item was designed to address the teacher's beliefs about his or her ability to incorporate technology into teaching, and another item addressed the teacher's beliefs about the use of technology and its relationship to student achievement. Thus, the number of items appearing in the final survey would be greater than the number of attributes.

Item wording addressed two main points, the first being what teachers believed about their ability to teach the specified task embedded in the standards-based attribute. The second point was what teachers believed about the performance of the task and its relationship to student achievement. These two parameters are the crux of the instrument's design.

A set of items was generated that referenced each of the standards-based attributes, half with positive-wording orientation, and half with negative-wording orientation. Attention was also given to a balanced number of items related to teacher self-efficacy beliefs and to beliefs about their relationship to student achievement. Word-smithing of each item ensured that language was as direct and unambiguous as possible. For example, the word "possible" conveys the idea that something might happen, or it might not. The words "likely" and "is" imply a higher probability of the event occurring. In the same way, the word "can" was used in favor of the word "may." Particular care was also taken to avoid compound sentence structure within each item. The examples in the following paragraphs illustrate how the wording reflects the design parameters.

As examples of the first point, what teachers believe about their ability related to the standards-based attribute, the following two items focus on designing experiments and investigations--one of the attributes. The difference between them, however, is that one refers to the teacher's beliefs about guiding students in designing their own experiments and investigations, and the other refers to the teacher's beliefs about designing the experiments and investigations for the students. The *National Science Education Standards* (NRC, 1996) are clear about the former task requiring a different type of teacher performance than the latter.

22. I am effective in guiding students to design their own experiments and investigations.	18. I am not very effective in developing science experiments and investigations for students.

The second point mentioned as a design parameter was what teachers believed about their performance of the task and its relationship to student achievement. The following pair of items relate to the teacher's beliefs about the relationship between planning and student achievement. One item expresses the belief that more time spent on planning has a positive effect on student achievement. The other emphasizes the belief that planning and using lessons that are connected and in sequence has a positive influence on student achievement.

6. If I spend more time planning science lessons, students' achievement will improve.	32. When I plan and use connected, sequenced lessons, my students' achievement in science improves.

A third pair of items is focused on the use of technology in science teaching, another standards-based attribute. In this pair of items, one addresses the teacher's self-efficacy belief about the use of technology, and the other addresses the teacher's belief about how the use of technology is related to student achievement.

41. I am comfortable using a variety of technologies in my science teaching.	44. When I use technology in my science class, it ensures that students will have higher achievement.

A Likert-type response scale was included with each item. The 5-point scale consisted of Strongly Agree (SA), Agree (A), Neutral (N), Disagree (D), and Strongly Disagree (SD). The highest value (5) was assigned to the response most consistent with positive implementation of the standards-based attribute, and the lowest value (1) was assigned to the response least consistent with the standards-based attribute.

Subjects and Data Collection

Participants for the study were recruited through established research protocols at the district level or through building principals if no such district-level protocols existed. Some survey packets were delivered in person to teachers, others were delivered by the district supervisor or department chair, and still others were sent by postal service after the initial contact with district- or school-level administrators. The information packet for each teacher included the following:

1. A single-page description of the study including confidentiality guidelines, design parameters, and contact information for the researcher--this page remained with the participant. Information was also included about the amount of time needed to complete the survey, approximately 25 minutes.
2. A participant consent form with a tear-off portion containing the researcher's contact information.
3. The survey, including information in the heading about the purpose of the study.

Surveys were distributed to teachers in districts across the state, and returns came from teachers in all areas of the state. They returned from schools in rural Appalachia and a large urban district; from schools with over 1,000 students and small schools with fewer than 300; and from a large suburban district with few families receiving financial assistance to inner city schools with more than 60% receiving aid. The surveys were returned by postal service or collected by the researcher in person, and the number of usable surveys totaled 103, which were used for the principal component analysis and for determining the reliability. After all missing data were accounted for the number of individuals with complete data was 87.

The demographic information including requests for name, year of birth, date, most advanced degree completed, year of completion of most advanced degree, number of years as a classroom teacher, and number of years teaching science, used a fill-in response format. A check box for marking a "Yes" response was used for other demographic questions. For example, under the subheading of "Degree Major," each subject area (e.g., Mathematics, Reading/Literacy, Science, Social Studies) was listed separately with its own accompanying check box. Responses for those items were then coded as a "0" for "No" and a "1" for "Yes."

Information was also collected about respondents' schools. School type was listed as public, private, or other--charter schools are an example of schools that would have been labeled "Other." Participants had the choice of labeling their school urban, suburban, or rural. The label "suburban" included schools in the outskirts of the city as well as those in areas farther away from the city center, but not into the agricultural areas. School size and socio-economic parameters were included on the survey to facilitate the respondents' designation of their schools and to offer information that is more descriptive about emerging relationships between variables.

Teacher preservice information showed the number and percent of teachers whose initial certification or licensure was earned in a 4-year undergraduate program or in a graduate-level program. Choices related to their major area of study (e.g., Mathematics, Reading/Literacy, Science, Social Studies) were offered with individual checkboxes so that the analyses might provide information that is more specific. The grades of initial certification or licensure offered choices in the same way. Grade ranges of the options represented types and levels of certification and licensure that have been available over the past 30 years in this state: PreK-8, K-8, 1-8, 4-9, 7-12, 9-12, and Other. It also represented the most common levels of certification in several other states (Council of Chief State School Officers, 2002). The final information in the subsection about preservice experiences represented the number and percent of respondents who had participated or presented at science or science education conferences at any level--local/district/state, national, or international.

The next set of demographic information collected related to the respondents' most advanced academic degree and the approximate number of graduate-level credits earned after the most advanced degree. A blank space was left for respondents to enter the degree name or initials, but numbers of graduate hours after the degree were clustered, and a checkbox was provided for each level: 0, 1-5, 6-10, 11-15, more than 15.

Information collected about participants' involvement in professional enhancement activities (e.g., classes, conferences) was collected in the next section.

Analytic Methodologies

A complete list of the independent variables is located in Table 3.3. The analysis began with the teacher and school demographic data and included several variations of frequency distributions. For example, for the teachers' schools, not only was an analysis made of each of the school characteristics independently, but also when compared to the other characteristics. The numbers of urban, suburban, and rural schools were counted, as were the numbers of small, medium, and large schools, and the numbers of schools with low, medium, and high percentages of families receiving financial aid. These numbers were examined within each grouping first, but were then analyzed across groups, resulting in information such as the number of large, urban schools with a high percentage of families receiving aid. Cross-group comparisons were made with three of the four categories of school demographics.

Frequency distributions were also calculated for gender, most advanced degree completed, and number of years' teaching--both career and science. The data for teacher preparation and professional development after initial certification were analyzed by each individual checkbox within the headings. Rather than ask participants to enter the information in a more open format, the checkboxes enabled a more detailed examination of the responses by creating a "Yes/No" type of choice for each item.

Independent Variables	
School Characteristics (3)	School Location School Size School SES
Teacher Characteristics (5)	Gender Most Advanced Degree Number of Years Teaching: Career Number of Years Teaching Science Approx. Number of Graduate Hours After Most Advanced Degree
Preservice Background (7)	Major in Science: Pre-Certification/Licensure Major-Other Type of Initial Certificate or License Preservice Teacher Program: Undergraduate or Graduate Level Quarters, Semesters, Mixed Attend Science Conference Pre-Certification/Licensure Attend Science Ed Conference Pre-Certification/Licensure
Science Teaching Experience (8)	Taught Biology Taught Chemistry Taught Physics Taught Physical Science Taught Earth Science Taught Integrated Science Taught General Science Taught Other Science
Post-Certification Professional Development (10)	University Science Class Participant University Science Ed Class Participant Continuing Ed Unit Science Participant CEU Science Ed Participant Local Science Conference Presenter Local Science Conference Participant Local Science Ed Conference Presenter Local Science Ed Conference Participant National Science Conference Participant National Science Ed Conference Participant

Table 3.3: Independent variables.

CHAPTER 4

RESULTS

Introduction

The primary purpose of this study was to create a valid, reliable instrument to further refine the measurement of self-efficacy among science teachers. As laid out in Chapter 1, the intent behind the creation of such a tool is to draw from it to influence science teacher education programs. The results presented in this chapter, and the interpretations in Chapter 5, may enlighten these possibilities.

Demographic information was gathered from the participants--48 questions about their schools, their teacher preparation program, and their professional development after initial certification or licensure were described in Chapter 3. Several of the dichotomous coded variables were not evenly distributed between the two categories. For example, in response to the question about school type, the vast majority of schools were public (96.1%) schools, not private (3.9%). The decision was made to eliminate demographic variables for which representation of either category was less than 13%.

Consequently, the first item to be eliminated described school type--public, private, or other. Only four teachers indicated that their school was anything other than public, so with such a small sample, the variable was set aside until another study. The next variables to be culled from the list included questions related to following: (a) Year of Most Advanced Degree, (b) Major in Math, (c) Major in Reading/Literacy, (d) Major in Social Studies, (e) University Science Class Presenter, (f) University Science Education Class Presenter, (g) Continuing Education Unit Science Presenter, (h) Continuing Education Unit Science Education Presenter, (i) National

Science Conference Presenter, (j) National Science Education Conference Presenter, (k) International Science Conference Presenter, (l) International Science Conference Participant, (m) International Science Education Conference Presenter, and (n) International Science Education Conference Participant. An examination of the returned surveys indicated that either participants did not bother to respond to these questions, or there were too few "yes" responses to provide useful information. This process resulted in 33 of the 48 independent or demographic variables being retained for subsequent analyses.

Principal Component Analysis

The first set of results in the principal component analysis showed a scree plot with a somewhat ambiguous display (See Figure 4.1). The optimum number of principal components was not immediately apparent--two components were obvious, but there was some question about the possibility of multiple additional components. Using SPSS, seven separate solutions of the principal component analysis were completed, each with one fewer components requested. The first solution showed loadings on 8 components, the second with 7, the third with 6, and so on, until the data were showing the loadings for 2 components. A 6-component solution did not converge through 25 iterations, however. Note that of the 60 items, #15, #19, #40, and # 42 are worded with a negative orientation. Their coding for the analyses was such that the highest score value was assigned to the response level with the strongest disagreement, thus a high level of disagreement was the most desired response.

Scree Plot

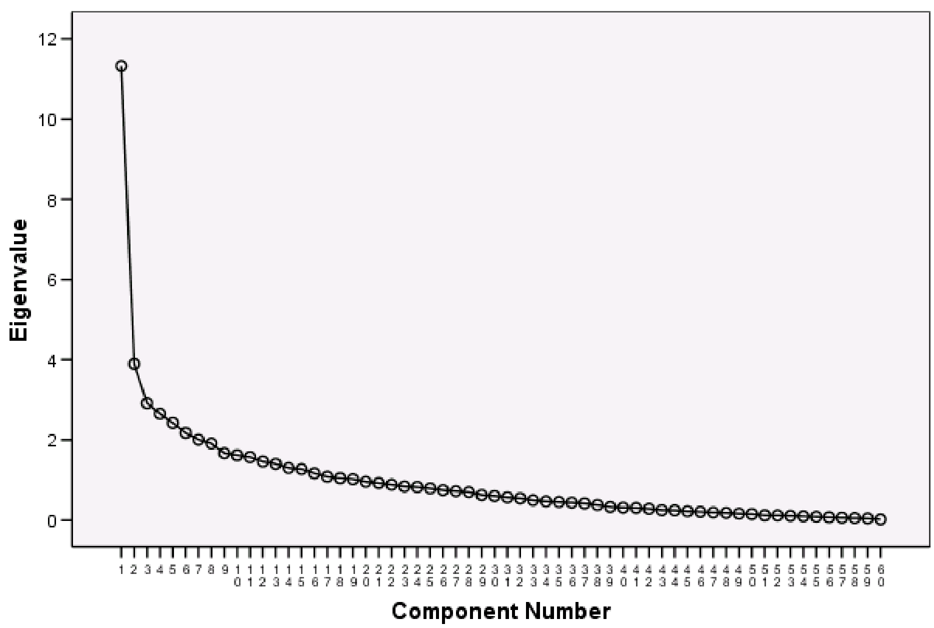


Figure 4.1: Scree plot of the principal component analysis for the Standards-Based Science Teacher Efficacy Instrument

The criteria used to determine the best solution were (a) the scree plot, (b) simple structure, and (c) interpretability. Items would be omitted from further analyses because of either low loadings or not fitting in the strongest final 2-component solution. Table 4.1 is a display of the loading values for both Component 1 and Component 2.

Items with Loading Values on Component 1 and Component 2

		n = 102	
		C 1	**C 2**
1.	For me, managing the classroom during a hands-on science activity is overwhelming.		**.35**
2.	I feel capable of teaching science to all students.	**.68**	
3.	I find it difficult to explain to students why science experiments and investigations come out the way they do.	**.41**	**.31**
4.	For me, issues of safety in the science classroom are not intimidating.	**.49**	
5.	I welcome students' science questions.	**.53**	
6.	If I spend more time planning science lessons, students' achievement will improve.		
7.	I understand science processes well enough to be effective in teaching science.	**.74**	
8.	Students' achievement in my science class is a direct result of my effectiveness in teaching science.		
9.	In my class, students' understanding of science concepts can be measured by how well they perform on traditional tests (e.g., multiple-choice, true/false, and matching.)		
10.	I believe I could teach a class that included advanced science concepts and processes.	**.62**	
11.	For me, the lecture is the most effective way to teach science concepts.		
12.	High achievement of students in my science class depends on memorizing definitions of scientific terminology.		
13.	I am comfortable teaching inquiry-based science lessons.	**.39**	**.56**
14.	Generally, I feel secure about teaching science.	**.77**	
15.	I don't know how to inspire students to do science.	**.64**	
16.	Every student in my class should have a science textbook to ensure high achievement.		
17.	When the science grades of students improve, it is often due to my having found a more effective teaching approach.		**.33**
18.	I am not very effective in developing science experiments and investigations for students.	**.48**	
19.	When a student has difficulty understanding a science concept, I am usually at a loss as to how to help the student understand it better.	**.62**	
20.	I know how to teach science concepts effectively.	**.57**	
21.	Even if I try very hard, I do not teach science very well.	**.58**	

Continued

Table 4.1: Items with loadings on Component 1 and Component 2.

Table 4.1 continued

22. I am effective in guiding students to design their own experiments and investigations.		**.33**
23. I can overcome the inadequacies of students' science background by good teaching.		**.44**
24. I can involve students in inquiry-based science with teacher-led demonstrations.		
25. Typically, I am able to answer students' science questions.	**.57**	
26. Even if I increase my effort in teaching science, some students will show little change in achievement.		**.38**
27. I understand science concepts well enough to be effective in teaching science.	**.71**	
28. In my class, students who participate regularly in inquiry-based science experiments and investigations have higher achievement scores.		**.57**
29. When students are underachieving in science in my class, it can be due to my ineffective science teaching.		**.36**
30. When parents/guardians of my science students comment that their child is showing more interest in science at school, it is probably due to my performance as a science teacher.		**.38**
31. It makes me uncomfortable to deal with the idea of teaching science to all students.	**.42**	
32. When I plan and use connected, sequenced lessons, my students' achievement in science improves.		**.37**
33. Students' achievement will be poor when my science lessons do not include hands-on activities.		**.39**
34. In my class, traditional tests (e.g., multiple-choice, true/false, and matching) are the best way to measure students' achievement in science.		
35. I feel at ease when students are working in cooperative learning groups during science experiments and investigations.		**.53**
36. In my class, students' achievement in science will improve when they work in cooperative learning groups.		**.65**
37. Having adequate science equipment and materials directly results in higher achievement by students in my class.	**.39**	
38. When my low-achieving students progress in science, it is usually due to extra attention I give as their science teacher.		
39. Given a choice, I will not invite the principal or supervisor to evaluate my science teaching.	**.40**	
40. I am not confident of my ability to teach science effectively.	**.71**	
41. I am comfortable using a variety of technologies in my science teaching.	**.54**	
42. I wonder if I have the necessary skills to teach science.	**.77**	
43. I will continually find ways to teach science more effectively.	**.65**	
44. When I use technology in my science class, it ensures that students will have higher achievement.		**.50**

Continued

Table 4.1 continued

45. In my class, students' achievement improves when science instruction is inquiry-based.		**.73**
46. I am not comfortable using alternative assessments (e.g., creating graphics, writing open-response, or designing projects) in science.	**.53**	
47. Students' achievement improves when I use self-evaluation and reflection about my science teaching.		**.41**
48. My students' achievement will not improve unless the science class is highly structured.		**.33**
49. All students can achieve at high levels in my science class.		**.38**
50. Teaching with hands-on activities in my science class makes me uncomfortable.	**.41**	
51. Low-achieving students can be motivated to improve in my science class.	**.40**	**.54**
52. I am comfortable with parental involvement in my science class.		
53. My own background in science helps me teach science more effectively.	**.50**	
54. Students who follow safety requirements for science experiments and investigations achieve at higher levels in my class.		
55. In my class, learning concepts in depth is directly related to higher student achievement in science.		**.36**
56. I can teach more effectively if I rely on only the science textbook for lesson plans.		**.33**
57. I cannot be an effective science teacher without adequate science equipment and materials.		
58. In my class, students must learn scientific processes in order to improve their understanding of science concepts.		
59. I am comfortable using scientific terminology in my science teaching.	**.60**	
60. I can measure students' depth of understanding of science concepts by using alternative assessments (e.g., creating graphics, writing open-ended responses, or designing projects).	**.41**	**.43**

The final determination was to keep 24 items that loaded from .77 to .39 for Component 1 (accounting for 19% of the variance) and to keep 21 items loading from.73 to .33 for Component 2 (accounting for 6% of the variance). Fifteen items did not load on either component and were deleted from further analyses and set aside for future development. Four items loaded on both Components 1 and 2. Two of the items (#3 and #60) had somewhat similar loadings on both components, so the decision was to omit both. Two other items (#13 and #51) displayed higher loadings for Component 2, so the decision was to add these two items to Component 2.

Table 4.2 presents the 24 items that loaded on Component 1. The first 11 items with the highest loadings from .77 to .60 are exemplars of the first component related to both self-efficacy and standards-based science. Negativity notwithstanding, the statements address an assessment of confidence and competence in personal teaching skills, knowledge of both science processes and concepts, and address an implied knowledge of what it takes to be effective as a teacher of standards-based science. These characteristics are identified and supported in the National Science Teachers document pertaining to standards for science teacher education programs. A logical decision was to name Component 1 as *Standards-Based Science Teacher Efficacy: Beliefs About Teaching* (BAT). The remaining 13 items that loaded on the BAT scale also had strong loading values, but the common characteristics were not always as clear-cut as that for the first 11 items. A complete list of all the items, with their loadings on BAT, is located in Table 4.2.

Items Loading on Component 1 (n = 102)		
14.	Generally, I feel secure about teaching science.	**.77**
42.	I wonder if I have the necessary skills to teach science.	**.77**
07.	I understand science processes well enough to be effective in teaching science.	**.74**
40.	I am not confident of my ability to teach science effectively.	**.71**
27.	I understand science concepts well enough to be effective in teaching science.	**.71**
02.	I feel capable of teaching science to all students.	**.68**
43.	I will continually find ways to teach science more effectively.	**.65**
15.	I don't know how to inspire students to do science.	**.64**
10.	I believe I could teach a class that included advanced science concepts and processes.	**.62**
19.	When a student has difficulty understanding a science concept, I am usually at a loss as to how to help the student understand it better.	**.62**
59.	I am comfortable using scientific terminology in my science teaching.	**.60**
21.	Even if I try very hard, I do not teach science very well.	**.58**
20.	I know how to teach science concepts effectively.	**.57**
25.	Typically, I am able to answer students' science questions.	**.57**
41.	I am comfortable using a variety of technologies in my science teaching.	**.54**
05.	I welcome students' science questions.	**.53**
46.	I am not comfortable using alternative assessments (e.g., creating graphics, writing open-response, or designing projects) in science.	**.53**
53.	My own background in science helps me teach science more effectively.	**.50**
04.	For me, issues of safety in the science classroom are not intimidating.	**.49**
18.	I am not very effective in developing science experiments and investigations for students.	**.48**
31.	It makes me uncomfortable to deal with the idea of teaching science to all students.	**.42**
50.	Teaching with hands-on activities in my science class makes me uncomfortable.	**.41**
39.	Given a choice, I will not invite the principal or supervisor to evaluate my science teaching.	**.40**
37.	Having adequate science equipment and materials directly results in higher achievement by students in my class.	**.39**

Table 4.2: Items loading on Standards-Based Science Teacher Efficacy: Beliefs About Teaching (BAT).

The second intention of the survey, after the first of measuring science teacher self-efficacy, was to attempt to measure science teacher efficacy beliefs about their role in their students' success in standards-based science classes. Of the first seven items with loadings ranging from .73 to .50 on the second component, five make direct reference to achievement in science. There are also direct references made to two important attributes of standards-based science: inquiry-based teaching and learning and working in collaborative groups. Three of the seven items (#45, #28, and #13) are tied to the former, two (#36 and #35) to the latter. Of the remaining 14 items, nine also have language connected to the outcome of student achievement, but their loadings were not as high as the first seven items. This second component was named *Standards-Based Science Teacher Efficacy: Beliefs About Student Achievement* (BASA), and a complete listing of the 21 items loading on BASA and their respective loading values is located in Table 4.3.

Items Loading on Component 2 (*n* = 102)		
45.	In my class, students' achievement improves when science instruction is inquiry-based.	**.73**
36.	In my class, students' achievement in science will improve when they work in cooperative learning groups.	**.65**
28.	In my class, students who participate regularly in inquiry-based science experiments and investigations have higher achievement scores.	**.57**
13.	I am comfortable teaching inquiry-based science lessons.	**.56**
51.	Low-achieving students can be motivated to improve in my science class.	**.54**
35.	I feel at ease when students are working in cooperative learning groups during science experiments and investigations.	**.53**
44.	When I use technology in my science class, it ensures that students will have higher achievement.	**.50**
23.	I can overcome the inadequacies of students' science background by good teaching.	**.44**
47.	Students' achievement improves when I use self-evaluation and reflection about my science teaching.	**.41**
33.	Students' achievement will be poor when my science lessons do not include hands-on activities.	**.39**
49.	All students can achieve at high levels in my science class.	**.38**
30.	When parents/guardians of my science students comment that their child is showing more interest in science at school, it is probably due to my performance as a science teacher.	**.38**
26.	Even if I increase my effort in teaching science, some students will show little change in achievement.	**.38**
32.	When I plan and use connected, sequenced lessons, my students' achievement in science improves.	**.37**
55.	In my class, learning concepts in depth is directly related to higher student achievement in science.	**.36**
29.	When students are underachieving in science in my class, it can be due to my ineffective science teaching.	**.36**
01.	For me, managing the classroom during a hands-on science activity is overwhelming.	**.35**
56.	I can teach more effectively if I rely on only the science textbook for lesson plans.	**.33**
17.	When the science grades of students improve, it is often due to my having found a more effective teaching approach.	**.33**
48.	My students' achievement will not improve unless the science class is highly structured.	**.33**
22.	I am effective in guiding students to design their own experiments and investigations.	**.33**

Table 4.3: Items loading on Standards-Based Science Teacher Efficacy: Beliefs About Student Achievement (BASA).

Internal Consistency Reliability Analyses

Reliability statistics for the resulting subscales is encouraging, the loss of 15 items notwithstanding. After an initial analysis using the 24 items loading on BAT and the 21 items loading on BASA, the results indicated that Item 49 on the BASA had a corrected item total correlation of -.37. The item was removed from further analyses due to its negativity. In a range from 24 to 120 points for all the items in the BAT subscale, and with a neutral score of 72, the mean score of the respondents was 101.5 with a standard deviation of 10.3. The neutral score of the second subscale, BASA, was 60 in the range between 20 and 100. The scale statistics showed a mean of 73.7 and SD of 7.8. The overall internal consistency reliability of the two subscales, Cronbach's alpha, based on standardized items, was .92 for the 24 items in BAT and .82 for the 20 items in BASA. Reliability data for both subscales can be found in Table 4.4.

Internal Consistency Reliability (Cronbach's Alpha) for BAT and BASA of the Standards-Based Science Teacher Efficacy Beliefs Instrument

Variable	No. of Items	No. of Subjects	Cronbach's Alpha
Component 1: Standards-Based Science Teacher Efficacy: Beliefs About Teaching (BAT)	24	102	.92
Component 2: Standards-Based Science Teacher Efficacy: Beliefs About Student Achievement (BASA)	21	102	.79
Component 2: Standards-Based Science Teacher Efficacy: Beliefs About Student Achievement (BASA)	20	102	.82

Table 4.4: Internal consistency reliability for BAT and BASA of the Standards-Based Science Teacher Efficacy Beliefs Instrument.

Analysis of the Independent Variables

The survey contained questions that might offer some insight into the relationships, if they exist, between teachers' experiences and their sense of self-efficacy and how that relates to their student achievement. A tacit understanding existed that some of the queries in the survey might not produce any useful data in this study, but that they might lead to more questions worth investigating in future research.

A second decision-point related to the independent variables involved the intent of the entire study. Its purpose was to discover relationships between certain variables and teacher efficacy. Of the 103 surveys returned to the researcher, 15 had data missing from the demographics section, and one had data missing from the survey items. In all calculations involving the independent variables, therefore, the number of usable responses dropped to 87. The list of the 33 independent variables clustered by categories and retained for the calculations of correlations and regressions is located in Table 4.4 and includes the variable means and standard deviations.

Independent Variables			
		M	*SD*
School Characteristics (3)	School Location	1.83	.62
	School Size	2.37	.63
	School SES	1.74	.92
Teacher Characteristics (5)	Gender	1.41	.49
	Most Advanced Degree	1.78	.48
	Number of Years Teaching: Career	14.35	10.33
	Number of Years Teaching Science	12.19	9.81
	Approx. Number of Graduate Hours After Most Advanced Degree	3.83	1.47
Preservice Background (7)	Major in Science: Pre-Certification/Licensure	.89	.31
	Major-Other	0.13	0.33
	Type of Initial Certificate or License	1.66	.48
	Preservice Teacher Program: Undergraduate or Graduate Level	1.32	.50
	Quarters, Semesters, Mixed	1.40	1.21
	Attend Science Conference Pre-Certification/Licensure	.13	.34
	Attend Science Ed Conference Pre-Certification/Licensure	.14	.35
Science Teaching Experience (8)	Taught Biology	.57	.50
	Taught Chemistry	.38	.49
	Taught Physics	.29	.46
	Taught Physical Science	.53	.50
	Taught Earth Science	.46	.50
	Taught Integrated Science	.41	.50
	Taught General Science	.57	.50
	Taught Other Science	.16	.37
Post-Certification Professional Development (10)	University Science Class Participant	.78	.42
	University Science Ed Class Participant	.74	.44
	Continuing Ed Unit Science Participant	.42	.50
	CEU Science Ed Participant	.36	.48
	Local Science Conference Presenter	.13	.34
	Local Science Conference Participant	.67	.47
	Local Science Ed Conference Presenter	.14	.35
	Local Science Ed Conference Participant	.67	.47
	National Science Conference Participant	.32	.47
	National Science Ed Conference Participant	.29	.46

Table 4.5: Independent variables with means and standard deviations.

Correlational Analysis

A correlation matrix was produced in which the intercorrelations among the demographic independent variables and Standards-Based Science Teacher Efficacy: Beliefs About Teaching (BAT) and Standards-Based Science Teacher Efficacy: Beliefs About Student Achievement (BASA) scores were explored. Table 4.6 displays correlations and levels of significance between BAT and BASA and each of the independent variables. The significant correlation between the two components is moderate (r = .44, p = .000) and implies that teachers with a positive sense of self-efficacy about teaching standards-based science are likely to believe that they can lead their students to achieve at higher levels in learning standards-based science. The relationships between pre-certification majors and both BAT and BASA subscales are worth noting--for both the correlation they show and what they do not show. For example, the entries for Major in Science show that teachers' having a pre-certification major in science is a poor predictor for a sense of efficacy in teaching standards-based science (r = .20, p = .066) and is not a predictor at all for beliefs about influencing student achievement in standards-based science (r = .00, p = .999). The next variable, Major-Other, however, did show significant negative correlations with both BAT and BASA (r =.-20, p = .028 and r = -.32, p = .003, respectively). It should be noted that the term "Major-Other" refers to any pre-certification major other than Mathematics, Reading/Literacy, Science, or Social Studies. Some other interesting significant positive correlations with both the BAT and BASA include CEU Science Education participant (r = .24, p = .025; r = .22, p = .040); Local Science Conference Participant (r = .26, p = .014; r = .24, p = .027); Local Science Education Conference Presenter (r = .28, p = .009); r = .26, p = .014); and Local Science Education Conference Participant (r = .33, p = .002; r = .34, p = .001). These results are not at all unexpected. Teachers who are more efficacious on both the BAT and BASA tend to participate in these types of professional development activities.

What is interesting is some of the differential results with regard to the BAT and BASA. For example, University Science Education Class Participant was positively correlated with the BAT (r = .29, p = .008) but not the BASA. In contrast, University Science Class Participant (r = .23, p = .033) at Local Science Conference Presenter (r = .22, p = .040) were positively correlated with BASA but not BAT. It appears that teacher who were more efficacious on the BAT tended to participate in science education professional development activities whereas teachers who were more efficacious on BASA tended to participate in science professional development activities.

Correlations

		BAT	BASA
Standards-Based Science Teacher Efficacy: Beliefs About Teaching (BAT)	*R*	1	
	P		
Standards-Based Science Teacher Efficacy: Beliefs About Student Achievement (BASA)	*R*	.44(***)	1
	P	0.000	
School Location	*R*	0.09	0.18
	P	0.408	0.097
School Size	*R*	-0.06	-0.17
	P	0.576	0.128
School SES	*R*	0.00	-0.03
	P	0.990	0.782
Gender	*R*	0.15	-0.07
	P	0.171	0.498
Preservice Teacher Program (Undergrad/Grad)	*R*	0.00	-0.09
	P	0.998	0.379
Quarters, Semesters, Mixed	*R*	0.03	0.15
	P	0.812	0.159
Major in Science	*R*	0.20	0.00
	P	0.066	0.999
Major - Other Not Math, Science, Reading/Literacy, SS	*R*	-.24(*)	-.32(**)
	P	0.028	0.003
Grades of Initial Certification	*R*	0.07	-.24(*)
	P	0.511	0.026
Most Advanced Degree	*R*	0.12	0.07
	P	0.253	0.507
Approx No of Grad Hrs Post Adv Deg	*R*	0.08	0.07
	P	0.459	0.552
Number of Yrs Teaching	*R*	0.10	0.09
	P	0.345	0.392
Number of Yrs Teaching Science	*R*	0.15	0.17
	P	0.180	0.111
Taught Biology	*R*	-0.02	-0.11
	P	0.894	0.328
Taught Chemistry	*R*	0.08	-0.08
	P	0.445	0.473

Continued

Table 4.6: Correlations between independent variables and BAT and BASA.

Table 4.6 continued

		BAT	BASA
Taught Physics	*R*	0.19	0.06
	P	0.073	0.609
Taught Physical Science	*R*	0.07	-0.07
	P	0.508	0.499
Taught Earth Science	*R*	-0.05	-0.15
	P	0.622	0.168
Taught Integrated Science	*R*	0.04	0.09
	P	0.705	0.419
Taught General Science	*R*	-0.04	-0.06
	P	0.726	0.576
Taught Other Science	*R*	0.15	-0.04
	P	0.151	0.713
University Science Class Participant	*R*	0.18	.23(*)
	P	0.103	0.033
University Science Ed Class Participant	*R*	.29(**)	0.19
	P	0.008	0.082
Continuing Ed Unit Science Participant	*R*	0.14	0.19
	P	0.194	0.083
CEU Science Ed Participant	*R*	.24(*)	.22(*)
	P	0.025	0.041
Local Science Conference Presenter	*R*	0.17	.22(*)
	P	0.119	0.040
Local Science Conference Participant	*R*	.26(*)	.24(*)
	P	0.014	0.027
Local Science Ed Conference Presenter	*R*	.28(**)	.26(*)
	P	0.009	0.014
Local Science Ed Conference Participant	*R*	.33(**)	.34(***)
	P	0.002	0.001
National Science Conference Participant	*R*	-0.04	0.06
	P	0.687	0.592
National Science Ed Conference Participant	*R*	0.13	0.20
	P	0.214	0.062

Note. * $p < .05$. ** $p < .01$. *** $p < .001$

As would be expected, certain demographics correlated very highly with each other. For example, number of years teaching science showed a significant positive correlation of *r* = .89 (*p* = 0.000) with its neighboring query--number of years teaching. Besides individual relationships between demographic variables, evidence emerged of clustering in certain areas of the matrix.

For example, the intersections of the two demographics about length of experience teaching and teaching science correlated with the six questions referring to science and science education conferences--both as presenter and as participant. It should be noted here that the term "science conference" refers to professional development opportunities offered at the local, district, state, or national level. The variable, Number of Years Teaching, correlated positively with Local Science Conference Presenter (*r* = .34, *p* = .001), Local Science Conference Participant (*r* = .27, *p* = .011), Local Science Education Conference Presenter (*r* = .40, *p* =.001), Local Science Education Conference Participant (*r* = .36, *p* = .001), National Science Conference Participant (*r* = .29, *p* = .006), and National Science Education Conference Participant (*r* = .37, *p* = .000). The variable, Number of Years Teaching Science, correlated positively with each of the same variables-- *r* = .28, *p* = .009; *r* = .34, *p* = .002; *r* = .32, *p* = .003; *r* = .39, *p* = .000; *r* = .34, *p* = .001; and *r* = .43, *p* = .000, respectively. This logical connection should not be surprising. Teachers who have been teaching longer, and those who have been teaching science longer, are more likely to have attended and/or presented at conferences.

In the examination of the correlations, there might be an expectations that most, if not all, the values would be positive. The data, however, produced several combinations that correlate inversely. For example, the teachers' type of certification shows a correlation of -.31 (*p* = .010) to participation in local science conferences, including school and district level sessions. The certification types were assigned nominal codes. Any certificate or license configuration for Pre-K through grade 9 was a value of 1, configurations for grades 7 through 12 were coded as a of 2, and teachers who had any other configuration were coded as 3. The type of certification also shows an inverse relationship to teaching general science--a teacher with a secondary teaching certificate is less likely to have taught general science than a teacher with a K-8 or 1-8 certificate or license ($r = -.41$, $p = .010$). Several interesting relationships involving some of the same variables emerge in the next section of the analysis--linear regression.

Multiple Stepwise Linear Regression Analysis

The next portion of the analyses involved performing a multiple stepwise linear regression using selected independent variables and each of the values for the efficacy subscales of Standards-Based Science Teacher Efficacy: Beliefs About Teaching (BAT) and Standards-Based Science Teacher Efficacy: Beliefs About Student Achievement (BASA). The goal was to attempt to identify which variables, if any, might be good predictors of teacher self-efficacy beliefs related to teaching standards-based science and facilitating standards-based student achievement.

From the regression analysis, three predictors accounting for 18% of the variance emerged for Component 1: Beliefs About Teaching-- Local Science Education Conference Participation, Major in Science, and Local Science Education Conference Presenter. The beta values (standardized coefficients) were 0.28, 0.26, and 0.22, respectively (see Table 4.6). The relationship of the trio of predictors is displayed with two models: the visual/conceptual diagram in Figure 4.2 and mathematically in the following equation where x_1 = Local Science Education Conference Participant; x_2 = Major in Science; and x_3 = Local Science Education Conference Presenter in standardized score form [$BAT = .28(x_1) + .26(x_2) + .22(x_3)$]. A discussion of the interconnections of the regression coefficients and their probable function in the equations for both BAT and BASA follows in Chapter 5. See Figure 4.2 for a visual display of the relationship of the predictors to BAT.

Coefficients

Dependent Variable: Standards-Based Science Teacher Efficacy: Beliefs About Teaching (BAT)

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta
1	(Constant)	96.97	1.78	
	Local Science Education Conference Participant	7.23	2.22	0.33
2	(Constant)	89.56	3.64	
	Local Science Education Conference Participant	7.71	2.18	0.36
	Major in Science	7.92	3.42	0.23
3	(Constant)	88.77	3.59	
	Local Science Education Conference Participant	6.15	2.26	**0.28**
	Major in Science	8.76	3.38	**0.26**
	Local Science Ed Conference Presenter	6.00	2.88	**0.22**
			n = 87	Adjusted R^2 = .18

Dependent Variable: Standards-Based Science Teacher Efficacy: Beliefs About Student Achievement (BASA)

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta
1	(Constant)	70.39	1.33	
	Local Science Education Conference Participant	5.43	1.65	0.34
2	(Constant)	71.72	1.38	
	Local Science Education Conference Participant	4.63	1.63	0.29
	Major - Other	-5.90	2.26	-0.26
3	(Constant)	80.04	3.28	
	Local Science Education Conference Participant	4.41	1.57	**0.27**
	Major - Other	-7.92	2.30	**-0.35**
	School Size	-3.39	1.22	**-0.28**
			n = 87	Adjusted R^2 = .22

Table 4.7: Stepwise multiple linear regression and adjusted R^2 for BAT and BASA.

Linear Model: Predictors for Standards-Based Science Teacher Efficacy: Beliefs About Teaching

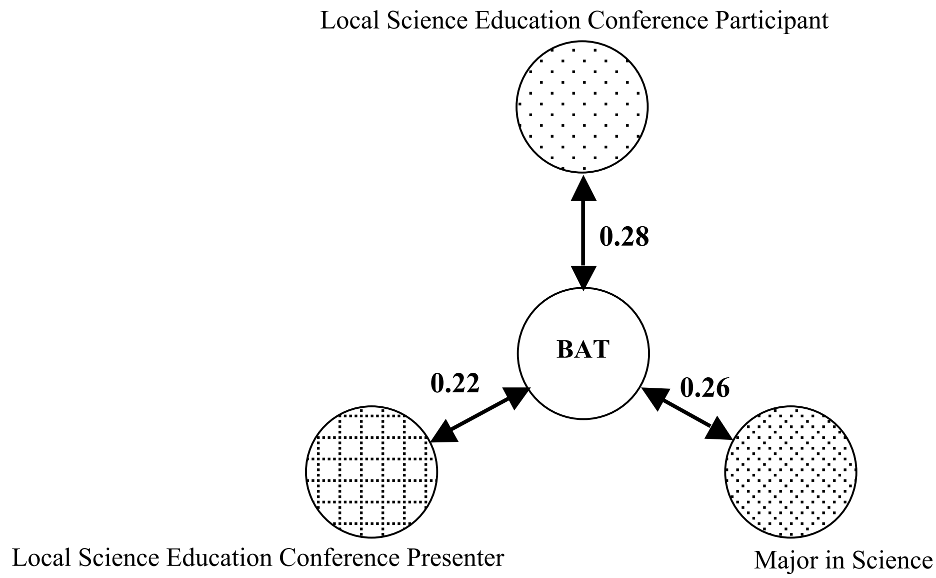


Figure 4.2: Linear regression model predictors for Standards-Based Science Teacher Efficacy: Beliefs About Teaching (BAT)

In the case of the predictors for science teacher efficacy beliefs about standards-based student achievement, interesting results emerged. As opposed to three positive values in the predictors for standards-based science teacher efficacy beliefs about teaching, this set resulted in one positive and two negative predictors, accounting for 22% of the variance. The positive predictor for BASA was the same variable as for BAT: Local Science Education Conference Participant (.27). The two negative predictors and their beta values were Major-Other (-.35) and School Size (-.28). The standardized score regression equation for this result is [$BASA = .27(x_1) + (-.35)(x_4) + (-.28)(x_5)$], where $x_1 =$ Local Science Education Conference Participant (the same predictor as in the first equation but with a different coefficient); $x_4 =$ Major-Other; and $x_5 =$ School Size. See Figure 4.3 for the graphic representation of the model of predictors for BASA.

Linear Model: Predictors for Standards-Based Science Teacher Efficacy: Beliefs About Student Achievement

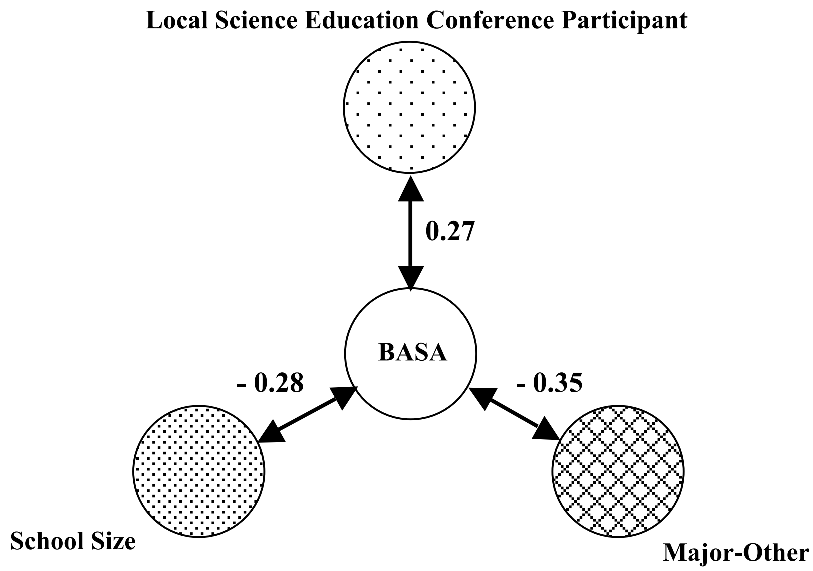


Figure 4.3: Linear regression model predictors for Standards-Based Science Teacher Efficacy: Beliefs About Student Achievement (BASA)

Further refinement and development of demographic variables and the BAT and BASA subscales will be discussed in Chapter 5 along with implications for future research. Omitted items and the crafting of new ones will be discussed along with a composite of the linear regression models for both BAT and BASA

CHAPTER 5

DISCUSSION AND IMPLICATIONS

Introduction

Creating a new instrument to measure science teacher self-efficacy beliefs and their beliefs about student achievement is not a simple task. The complexity increases exponentially when the idea of an infrastructure of national science education standards is put into the fray. The first four chapters lay out the idea, the theoretical background, the plan and design, the results, and validation. This chapter encapsulates the interpretations and implications for each of the major aspects of the study.

In the cogitation of results, and in light of expounding on the merits of a logic model, a first step is to return to the original problem statements and research questions, which can be interpreted as goals of this study. They will be revisited here and each will be discussed in light of the results. The main, underlying purpose of the study was to create a viable instrument. Not only was the intended instrument to address national science education standards, but also to reference standards for science teacher education programs, and determine the relationships between teacher beliefs about teaching standards-based science and their beliefs about student achievement in the same context.

Demographic Variables: Discussion and Future Research

With the thought of creating an instrument to measure science teacher efficacy beliefs about both standards-based teaching and standards-based student achievement, the thoughtful design of demographic questions and items was crucial. For teacher demographics, a parameter of the design was to differentiate between teacher experiences both before initial certification or licensure and after certification/licensure. The original set of seven reflective questions led, eventually, to the development of 48 questions about schools, professional development experiences--both preservice and inservice, type of professional certification/licensure, and conference attendance as presenter or participant. Having such a large number of questions may have influenced the responses. Insufficient data or an extreme imbalance between the two categories in the dichotomous coding pared the list of 48 demographic questions to 33. These questions became the list of independent variables for computation of the correlations and multiple linear stepwise regression analyses. Although shortening response time, eliminating 15 queries from the final round of analyses may leave a large gap in the information derived. There are demographic questions, such as the current grade level and subject teaching assignment, that could provide additional insights into the relationships, but insufficient data were collected to make that determination.

One of the main concerns in future studies is the identification and selection of appropriate independent variables. Suggestions for further research include categorizing demographic questions with more efficient coding, determining which are most viable as data collection questions, and eliminating questions that provided little or no relevant information.

Components: Discussion and Future Research

The 60 items developed for the survey supported the design by loading on two components in the principal components factor analysis-Standards-Based Science Teacher Efficacy: Beliefs About Teaching (BAT) and Standards-Based Science Teacher Efficacy: Beliefs About Student Achievement (BASA).

One of the most striking aspects of the results is that the respondents seem to have a fairly strong sense of efficacy about their teaching being aligned with standards, but their beliefs about student achievement in the same context is weaker. Teachers' beliefs about their own self-efficacy are definitely related to their beliefs about student achievement, and teachers seem to be well aware of the relationship. The difficulty is much the same as the big idea and the details about teaching. They acknowledge the relationship but falter somewhat in their own performance of tasks to accomplish the success. An interpretation might be that they believe in themselves as teachers doing an effective job, but the students are not learning at the same level of accomplishment and there is only so much a teacher, this teacher, can do.

Another seeming discrepancy is in the inclusion of assessments that are more authentic than traditional types of tests. Responses showed that teachers believed they could include authentic assessments but that their students would not be very successful with them. An opposing bit of information, but along the same lines, is the adherence to using lecture, a tightly structured classroom, and mandatory textbooks--all very traditional.

There is also a gap in the idea of teacher collegiality, parental involvement, and reflexive activities being assets to a standards-based program and related to strengthening teacher efficacy beliefs. Another aspect is that the respondents reacted positively to items that contained language using the terms “inquiry-based” and “collaborative learning,” but they responded more negatively to items that contained references to specific activities that comprise both “inquiry-based” and “collaborative learning.” A deeper study of the relationship between teacher efficacy beliefs about using “inquiry-based” activities and collaborative learning in standards-based science classrooms, and the specific activities and strategies that constitute those two ideas, is needed to clarify questions surrounding the items.

Another dichotomy is in reference to the use of technology in the classroom. It is definitely a characteristic of science teaching and learning espoused by the standards, but there are so many extenuating circumstances affecting the incorporation of technology that it needs another approach in efficacy studies. Teachers generally felt efficacious about their ability to include uses of technology, but the responses were not clear-cut about its role in student success.

Teachers' self-efficacy beliefs about their own science teaching are generally positive. Since no items referred to beliefs about teaching in general, the interpretation needs to be done carefully. They seem to have the big idea of what it means to have a standards-based science curriculum and to do inquiry-based science; however, they are much less confident when presented with specific tasks that constitute the bigger idea. For example, interactive lessons are important but items about students generating their own experiments and investigations, or the teacher guiding such activities, did not have a strong showing.

The items loading on Component 2-Standards-Based Science Teacher Efficacy: Beliefs About Student Achievement (BASA)--at first glance are closely aligned with standards-based language, but a deeper examination of the items not loading on Teacher Beliefs about Standards-Based Student Achievement reveals that there is another telling omission. Several items dealing specifically with tasks that are grounded in the standards, such as improving student achievement with improved planning of interactive science lessons, dealing with issues related to equity and diversity in the science classroom, and maintaining the materials and equipment used in an interactive science curriculum were not in the final discussions of Component 2.

Instrument Items: Discussion and Future Research

Development parameters, such as maintaining a balance between items with positive- and negative-wording orientation, balancing items referring to standards-based science teacher beliefs and teacher beliefs about standards-based student achievement, and cross-balancing all four issues were critical in the development of a validated instrument. Following the principal components factor analysis, 15 of the original items had been eliminated due to loadings below .30 or not loading on either of the two principal components, leaving 24 items that loaded on BAT and 21 items loading on BASA. A calculation of the reliability showed a negative discrimination value for one of the items on the BASA, eliminating it from further analyses. A repetition of the reliability calculations with 24 items loading on BAT and 20 items loading on

BASA resulted in Cronbach's alpha of .92 and .82, respectively, again supporting the item design.

Additional studies are definitely needed, because several items need to be rewritten to dispel ambiguity, clarify uncertainties, or simplify complexities. For example, Item 9 reads "In my class, students' understanding of science concepts can be measured by how well they perform on traditional tests (e.g., multiple-choice, true/false, and matching.)" This item did not load on either Component 1 or Component 2 of the final principal component analysis.

Other important items were omitted from the final analyses because of low loadings, loadings on both factors, or loadings on other factors. Of special interest were five items all including language and ideas more closely connected to pre-standards, traditional practices such as lecture and memorizing terminology, and the necessity of each and every student to have a textbook. This seems to add to the evidence of the complexity of the entire endeavor.

Items omitted after the principal component factor analysis and after the reliability calculations, should be examined for commonalities, especially if they loaded on a third or fourth component. It would provide insight into other aspects of standards-based science teacher efficacy beliefs that were not as fully addressed in this study. For example, items relating to technology should be expanded considering its role in modern society and culture.

Predictor Variables: Discussion and Future Research

As a result of stepwise multiple linear regression analysis to identify a set of predictor variables from the independent variables included in this study, two regression models were produced. The first model for predicting Standards-Based Science Teacher Efficacy: Beliefs About Teaching (BAT) included a set of three significant contributing variables. These three variables represent the general categories of professional development and science content preparation. The second model for predicting Standards-Based Science Teacher Efficacy: Beliefs About Student Achievement (BASA) also included three significant contributing variables which included the professional development and content preparation categories as well as the general category of teaching environment. The best single predictor in each case was Local Science Education Conference Participant. In addition to the sets of predictions identified by the regression analysis, it should also be noted that a significant correlation was found between BAT and BASA. The contribution of the interrelationships between the efficacy scales and their sets of predictions is provided as a schematic in Figure 5.1.

Results of the correlation analysis also reveal some interesting relationships that warrant further investigation. Some of the science education professional development activities were related to BAT but not BASA; similarly, some of the science professional development activities were related to BASA but not BAT.

Local and state level professional development activities seem to have a strong positive connection to standards-based science teacher efficacy beliefs and beliefs about student achievement. These experiences, coupled with a science content background seem to characterize high levels of science teacher efficacy.

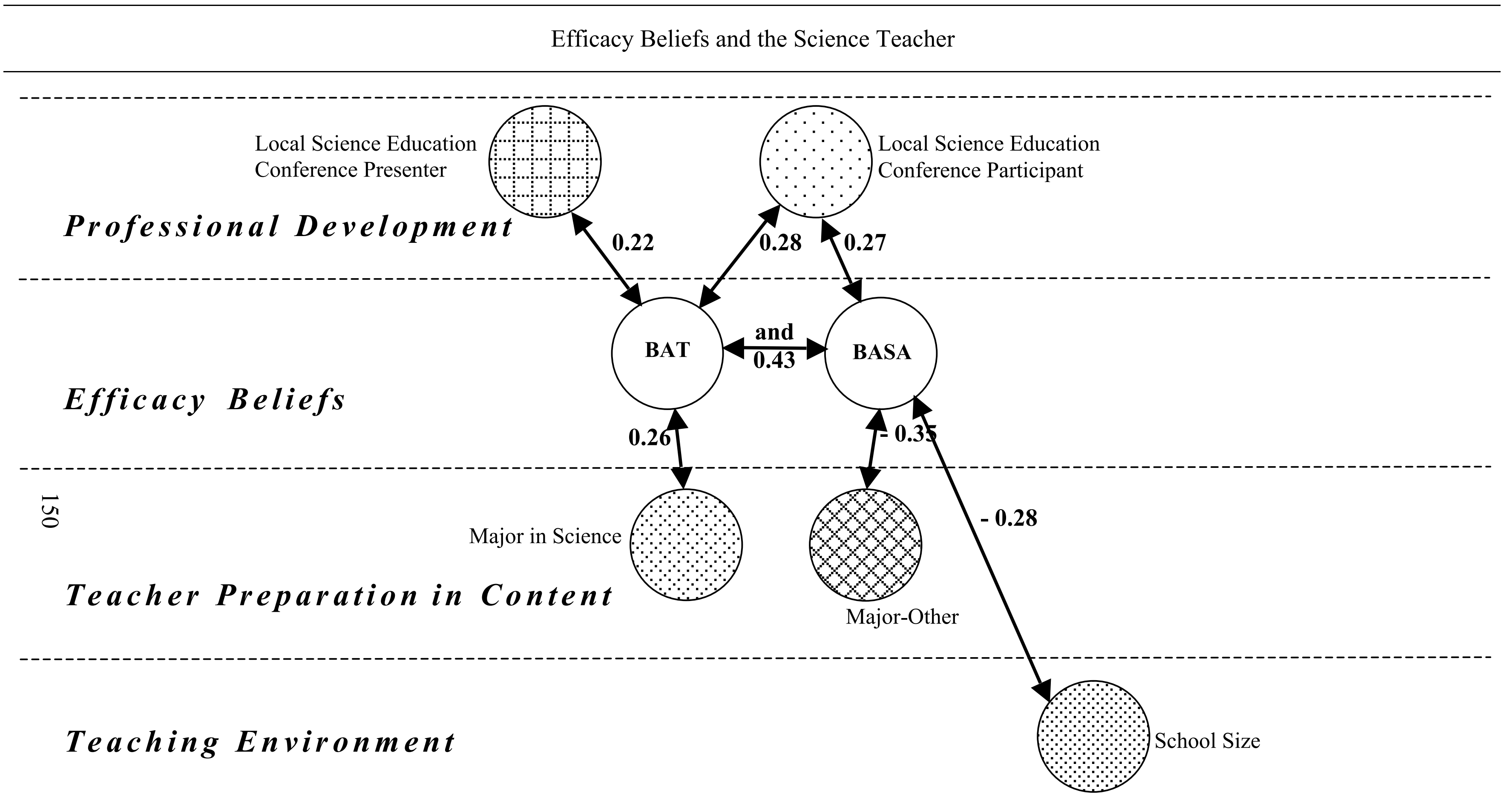


Figure 5.1: Efficacy beliefs and the science teacher

Connecting to the Literature

Several of the issues raised by this study and enlightened by the data analyses align with research literature in the progression of the study of teacher efficacy. For example, the self-efficacy beliefs relative to personal knowledge and personal teaching capabilities appear strong, while beliefs in connecting those capabilities to student achievement are less so. This is in line with previously cited research that states the same idea (Armor et al., 1976, Dembo & Gibson, 1985). Another issue is the connection between conference participation and science teacher efficacy. The relationship is bidirectional, but definite. Science teacher education program standards (NSTA, 2003) and documents developed as guides for quality professional development for science teachers support the connection (Nelson, 1997). The general trends of the findings align with, and provide additional support for, consistent research-based conclusions and recommendations.

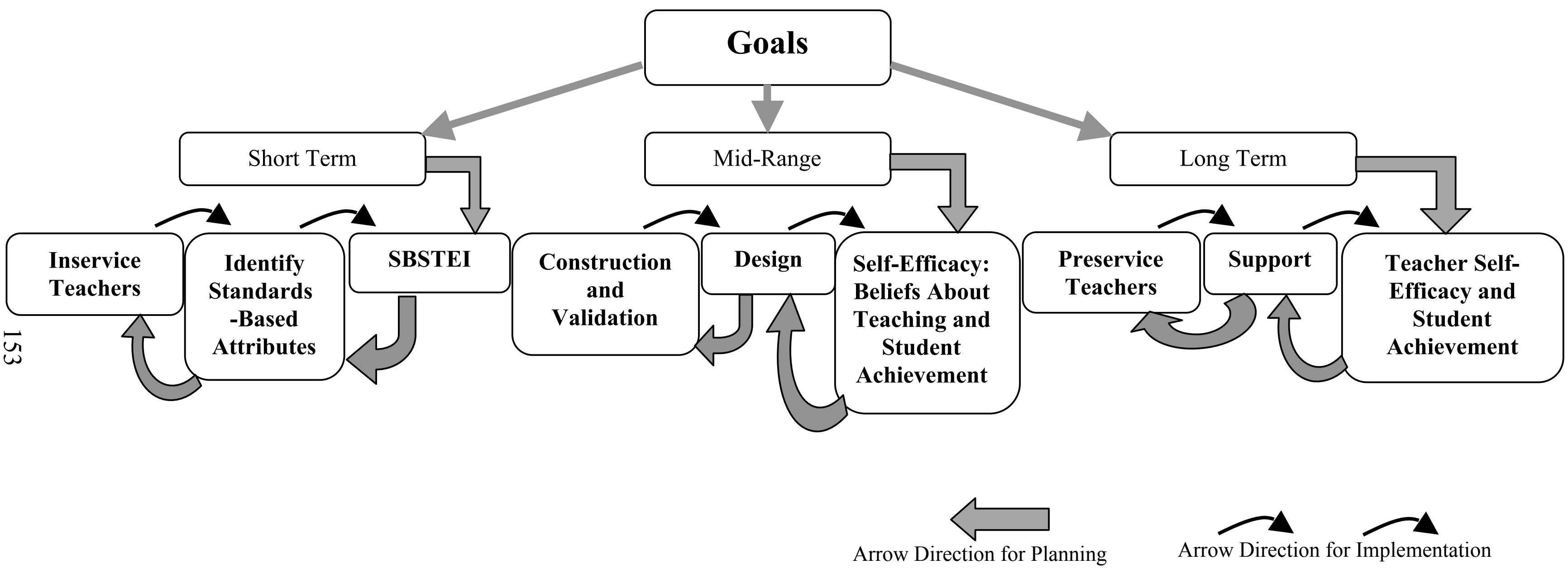
Logic Model-Next Steps

The following schematic, Figure 5.2, shows the application of a logic model to future goals and further development of the Standards-Based Science Teacher Efficacy Beliefs Instrument. The development and validation of this instrument is one component of a mid-range goal related to strengthening science teacher self-efficacy beliefs and their beliefs about student achievement.

The teacher characteristics of continually striving for professional self-improvement and being proactive about improving teaching techniques to promote student achievement indicate a recursive process involving self-reflection, self-evaluation, and self-improvement. Striving toward well-defined goals also involves acquiring a target (metaphorically speaking), analyzing the skills and knowledge needed to meet the target, comparing personal skills and knowledge of the topic to the needed ones, and developing a plan to reach the target. This is an essential process for both inservice and preservice teachers and is strongly influenced by teacher efficacy beliefs.

Future research could focus upon the relationships between teacher efficacy beliefs, actual classroom practice, and student achievement.

Further Development and Validation of a Standards-Based Science Teacher Efficacy Instrument in a Logic Model Schematic



Note. SBSTEI = Standards-Based Science Teacher Efficacy Instrument.

Figure 5.2: Further development and validation of the Standards-Based Science Teacher Efficacy Instrument in a logic model schematic

REFERENCES

American Association for the Advancement of Science. (1989). *Project 2061: Science for all Americans.* New York: Oxford University Press.

American Association for the Advancement of Science. (1993). *Benchmarks for scientific literacy.* New York: Oxford University Press.

American Association of University Women. (1992). *How schools shortchange girls*. New York: Marlowe & Company.

Armor, D., Conroy-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., et al. (1976). *Analysis of the school preferred reading program in selected Los Angeles minority schools* (No. R-2007-LAUSD). Santa Monica, CA: The Rand Corporation.

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

Ashton, P.T. (1982). *A study of teachers' sense of efficacy (*Volume I*,* Final Report). Gainesville: Foundations of Education, University of Florida.

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

Ashton, P.T., & Webb, R.B. (1986). *Making a difference: Teachers' sense of efficacy and student acievement.*Unpublished manuscript, New York.

Bandura, A, (1971). *Social learning theory.* Morristown, NJ: Prentice-Hall.

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

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

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

Bandura, A. (2001). *Guide for constructing self-efficacy scales*. Retrieved October 13, 2001, from http://www.coe.ohio-state.edu/ahoy/researchinstruments.htm#Ban

Battle, E.S., & Rotter, J.B. (1963). Children's feelings of personal control as related to social class and ethnic group. *Journal of Personality, 31*(4), 482-490.

Bleicher, R.E. (2004). Revisiting the STEBI-B: Measuring self-efficacy in preservice elementary teachers. *School Science and Mathematics, 104*(8), 383-391.

Cannon, J.R. (1996). Influence of a cooperative early field experience on preservice elementary teachers' science self-efficacy. *Science Education, 80*(4), 419-436.

Cantrell, P. (2003). Traditional vs retrospective pretests for measuring science teaching efficacy beliefs in preservice teachers. *School Science and Mathematics, 103*(4), 177-185.

Council of Chief State School Officers. (2004). *Interstate New Teacher Assessment and Support Consortium (INTASC)*. Retrieved June 5, 2004, from http://www.ccsso.org/projects/Interstate_New_Teacher_Assessment_and_Support_Consortium/

Deemer, S.A., & Minke, K.M. (1999). An investigation of the factor structure of the teacher efficacy scale. *The Journal of Educational Research, 93*(1), 3-10.

Dembo, M.H., & Gibson, S. (1985). Teachers' sense of efficacy: An important factor in school improvement. *The Elementary School Journal, 86*, 173-184.

Dembo, M.H., & Gibson, S. (1986). Teachers' sense of efficacy. *The Education Digest, 51*, 35.

Enochs, L.G., Oliver, S., & Wright, E.L. (1990a). An evaluation of the perceived needs of secondary science teachers in Kansas. *Journal of Science Teacher Education, 1*(4), 74-79.

Enochs, L.G., & Riggs, I.M. (1990). Further development of an elementary science teaching efficacy belief instrument: A preservice elementary scale. *School Science and Mathematics, 90*(8), 694-706.

Enochs, L.G., Riggs, I.M., & Scharmann, L.C. (1995). The relationship of pupil control to preservice elementary science teachers' self-efficacy and outcome expectancy. *Science Education, 79*(1), 63-75.

Enochs, L.G., Smith, P.L., & Huinker, D. (2000). Establishing factorial validity of the mathematics teaching efficacy beliefs instrument. *School Science and Mathematics, 100*(4), 194-202.

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

Goddard, R.D., Hoy, W.K., & Woolfolk Hoy, A. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal, 37*(2), 479-507.

Guskey, T.R. (1988).Teacher efficacy, self concept, and attitudes, toward the implementation of instructional innovation. *Teaching and Teacher Education: An International Journal of Research and Studies*. *4*(1), 63-68.

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

Henson, R.K. (2002). From adolescent angst to adulthood: Substantive implications and measurement dilemmas in the development of teacher efficacy research. *Educational Psychologist, 37*(3), 137-150.

Hoy, W. K., & Woolfolk, A. (1993). Teachers' sense of efficacy and the organizational health of schools. *Elementary School Journal, 93*(4) 355-372.

Kerr, T.R. (2002). *Exploring preservice science teacher attitudes and beliefs about science and science teaching*. Unpublished manuscript.

Lawrenz, F.P. (1975). *The relationship between teacher characteristics and student achievement and attitude* (Research Paper No. 8). Minneapolis: University of Minnesota.

Lowrey, L.F. (Ed.). (2000). *Pathways to the science standards: Elementary school edition* (Second ed. Vol. 1). Arlington, VA: National Science Teachers Association.

Luft, J.A., & Patterson, N.C. (2002). Bridging the gap: Supporting beginning science teachers. *Journal of Science Teacher Education, 13*(4), 267-282.

Muijs, D., & Reynolds, D. (2002). Teachers' beliefs and behaviors: What really matters? *Journal of Classroom Interaction, 37*(2), 3-15.

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

National Research Council. (1996). *The national science education standards*. Washington, DC: National Academy Press.

National Science Teachers Association. (2003). *Standards for science teacher preparation*. Retrieved December 12, 2003, from http://www.nsta.org/main/pdfs/NSTAstandards2003.pdf

Nelson, G. (1997). *Benchmarks and standards as tools for science education reform*. Retrieved January 5, 2005, from www.project2061.org/publications/articles/nelson/nelson1.htm

Pajares, F. (1997). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*, 543-578.

Parker, M.J., Guarino, A.J., & Smith, R.W. (2002). Self-efficacy in a sample of education majors and teachers. *Psychological Reports, 91*(3), 935-939.

Plourde, L.A. (2002a). Elementary science education: The influence of student teaching--Where it all begins. *Education, 123*(2), 253-259, 245.

Plourde, L.A. (2002b). The Influence of student teaching on preservice elementary teachers' science self-efficacy and outcome expectancy beliefs. *Journal of Instructional Psychology, 29*(4), 245-253.

Posnanski, T.J. (2002). Professional development programs for elementary science teachers: An analysis of teacher self-efficacy beliefs and a professional development model. *Journal of Science Teacher Education, 13*(3), 189-220.

Raizen, S.A., & Michelsohn, A.M. (Eds.). (1994). *The future of science in elementary schools: Educating prospective teachers*. San Francisco: Jossey-Bass.

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

Ritter, J.M., Boone, W.J., & Rubba, P.A. (2001). Development of an instrument to assess prospective elementary teacher self-efficacy beliefs about equitable science teaching and learning (SEBEST). *Journal of Science Teacher Education, 12*(3), 175-198.

Roberts, J.K., Henson, R.K., & Tharp, B.Z. (2001). An examination of change in teacher self-efficacy beliefs in science education based on the duration of inservice activities. *Journal of Science Teacher Education, 12*(3), 199-231.

Roehrig, G.H., & Kruse, R.A. (2005). The role of teachers' beliefs and knowledge in the adoption of a reform-based curriculum. *School Science and Mathematics, 105*(8), 412-422.

Rose, J.S., & Medway, F.J. (1981). Measurement of teachers' beliefs in their control over student outcome. *Journal of Educational Research, 74*(3), 185-189.

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

Rotter, J.B. (1966). Generalized expectancies for internal versus external control of reinforcement. [Monograph]. *American Psychological Association, 80,* 1-28.

Rotter, J.B. (1982). *The development and applications of social learning theory: Selected papers*. New York: Praeger.

Simmons, P.E., Emory, A., Carter, T., Coker, T., Finnegan, B., Crockett, D., et al. (1999). Beginning teachers: Beliefs and classroom actions. *Journal of Research in Science Teaching, 36*(8), 930-954.

Soodak, L.C., & Podell, D.M. (1996). Teacher efficacy: Toward the understanding of a multi-faceted construct. *Teaching and Teacher Education, 12*(4), 401- 411.

Soodak, L.C., & Podell, D.M. (1997). Efficacy and experience: Perceptions of efficacy among preservice and practicing teachers. *Journal of Research and Development in Education, 30*(4), 214-221.

Tosun, T. (2000). The beliefs of preservice elementary teachers toward science and science teaching. *School Science and Mathematics, 100*(7), 374-379.

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

Wenner, G. (1993). Relationship between science knowledge levels and beliefs toward science instruction held by preservice elementary teachers. *Journal of Science Education and Technology, 2*(3), 461-468.

Wenner, G. (1995). Science knowledge and efficacy beliefs among preservice elementary teachers: A follow-up study. *Journal of Science Education and Technology, 4*(4), 307-315.

Wenner, G. (2001). Science and mathematics efficacy beliefs held by practicing and prospective teachers: A 5-year perspective. *Journal of Science Education and Technology, 10*(2), 181-187.

Wiggins, G., & McTighe, J. (1998). *Understanding by design*: Association for Supervision and Curriculum Development.

Willower, D.J., Eidell, & Hoy, W.K. (1967). *The school and pupil control ideology*. Unpublished manuscript, University Park, PA.

Woolfolk, A., & Hoy, W.K. (1990). Prospective teachers' sense of efficacy and beliefs about control. *Journal of Educational Psychology, 82*(1), 81-91.

APPENDIX A

PILOT STUDY: MIDDLE LEVEL SCIENCE TEACHING EFFICACY AND OUTCOME BELIEF INSTRUMENT

Name **Today's Date**

Date of Birth ______________________________ **Gender** __________________________________

Licensure Area (Circle one.) Grades Pre-K - 3 Grades 4 - 9 Grades 7-12

Area(s) of Concentration: Please circle appropriate subject(s).

Mathematics Reading/Language Arts Science Social Studies

If Science, please circle the subjects you expect to be licensed to teach.

Chemistry Earth Sciences Life Science Physics

Number of <u>content</u> courses completed up to now. Please circle the appropriate number.

Mathematics	0	1-5	6-10	11-15	More than 15
Science	0	1-5	6-10	11-15	More than 15
Social Studies	0	1-5	6-10	11-15	More than 15
Reading/Language Arts	0	1-5	6-10	11-15	More than 15

Number of <u>methods</u> classes completed up to now. Please circle the appropriate number.

Mathematics	0	1	2-4	5-7	More than 8
Science	0	1	2-4	5-7	More than 8
Social Studies	0	1	2-4	5-7	More than 8
Reading/Language Arts	0	1	2-4	5-7	More than 8

Are you a full-time or part-time student? (Please check one.) ___ Full-time ____ Part-time

College(s) or University(ies) attended **Current University**

Undergraduate ____________________________ **Undergraduate** ____________________________

Graduate _________________________________ **Graduate** _________________________________

Please indicate the level of agreement or disagreement with each statement below by circling the appropriate letters to the right of each statement.

Strongly Disagree **SD** Disagree **D** Neutral or Uncertain **N** Agree **A** Strongly Agree **SA**

01. Managing the classroom during a hands-on science activity is overwhelming.	SD	D	N	A	SA
02. I feel capable of teaching science to all students.	SD	D	N	A	SA
03. I find it difficult to explain to students why science experiments come out the way they do.	SD	D	N	A	SA
04. Issues of safety in the science classroom do not intimidate me.	SD	D	N	A	SA
05. During science class, I welcome student questions.	SD	D	N	A	SA
06. A student's achievement will improve if the teacher spends more time planning.	SD	D	N	A	SA
07. I understand science processes well enough to be effective in teaching science.	SD	D	N	A	SA
08. Students' achievement in science is a direct result of their teacher's effectiveness in science teaching.	SD	D	N	A	SA
09. A student's understanding of science concepts is best measured by how well s/he performs on objective tests.	SD	D	N	A	SA
10. I believe I could teach a class that included advanced science concepts and processes.	SD	D	N	A	SA
11. The lecture is the most effective way for me to teach science concepts.	SD	D	N	A	SA
12. High achievement in science depends heavily on learning lists of vocabulary words.	SD	D	N	A	SA
13. I am prepared to plan for inquiry-based science lessons.	SD	D	N	A	SA
14. Generally, I feel secure about teaching science.	SD	D	N	A	SA
15. I don't know how to turn students on to science.	SD	D	N	A	SA

Strongly Disagree **SD** Disagree **D** Neutral or Uncertain **N** Agree **A** Strongly Agree **SA**

16. Every student should have a science textbook to ensure high achievement.	SD	D	N	A	SA
17. When the science grades of students improve, it is often due to their teacher having found a more effective teaching approach.	SD	D	N	A	SA
18. I am not very effective in developing science experiments for students.	SD	D	N	A	SA
19. When a student has difficulty understanding a science concept, I am usually at a loss as to how to help the student understand it better.	SD	D	N	A	SA
20. I know how to teach science concepts effectively.	SD	D	N	A	SA
21. Even if I try very hard, I do not teach science very well.	SD	D	N	A	SA
22. I am effective in guiding students to design their own experiments.	SD	D	N	A	SA
23. The inadequacy of a student's science background can be overcome by good teaching.	SD	D	N	A	SA
24. The best way to involve students in science activities is through teacher-led demonstrations.	SD	D	N	A	SA
25. Typically, I am able to answer students' science questions.	SD	D	N	A	SA
26. Even if I increase my effort in teaching science, some students will show little change in achievement.	SD	D	N	A	SA
27. I understand science concepts well enough to be effective in teaching science.	SD	D	N	A	SA
28. The teacher is generally responsible for low achievement of students in science.	SD	D	N	A	SA
29. When students are underachieving in science, it is most likely due to ineffective science teaching.	SD	D	N	A	SA
30. When parents/guardians comment that their child is showing more interest in science at school, it is probably due to the performance of the child's teacher.	SD	D	N	A	SA

Strongly Disagree **SD** Disagree **D** Neutral or Uncertain **N** Agree **A** Strongly Agree **SA**

Circle your response.

31. It makes me uncomfortable to deal with the idea of teaching science to all students.	SD	D	N	A	SA
32. Student achievement in science will be poor when the teacher does not plan connected and sequenced lessons.	SD	D	N	A	SA
33. A student's achievement will be poor when the science lessons do not include hands-on activities.	SD	D	N	A	SA
34. When some students have low achievement in science, the teacher is responsible.	SD	D	N	A	SA
35. I feel at ease when students are doing group work in science experiments and investigations.	SD	D	N	A	SA
36. Student achievement in science will improve when students work in groups.	SD	D	N	A	SA
37. Having adequate materials for science does not automatically result in higher student achievement.	SD	D	N	A	SA
38. When a low-achieving student progresses in science, it is usually due to extra attention given by the teacher.	SD	D	N	A	SA
39. Given a choice, I will not invite the principal to evaluate my science teaching.	SD	D	N	A	SA
40. I am not confident of my ability to teach science effectively.	SD	D	N	A	SA
41. When a student does better than usual in science, it is often because the teacher exerts extra effort.	SD	D	N	A	SA
42. I wonder if I have the necessary skills to teach science.	SD	D	N	A	SA
43. I will continually find better ways to teach science.	SD	D	N	A	SA
44. Using technology in science class ensures that students will have higher achievement.	SD	D	N	A	SA

APPENDIX B

PARTICIPANT CONSENT FORM

PARTICIPANT CONSENT FORM

I, ________________________ of ______________________ School,

(Teacher's name; please print) *(Name of school; please print)*

agree to participate in the study, called **Science Teachers' Beliefs,** conducted by Tricia Reda Kerr of The Ohio State University.

The nature and general purpose of the research procedures have been explained to me. I understand that I will receive an explanation of the research and may decline to participate. Further, I understand that I may terminate participation in this research at any time desired by contacting Tricia Reda Kerr in writing at the address listed below.

Signed: ______________________________________Date: ________________________

Research Co-Investigator: _____Tricia R. Kerr_____________________

Tear along the dotted line to retain contact information for your records.

- -

Contact Address for Research: *Science Teachers' Beliefs*

Tricia Reda Kerr
c/o Dr. Donna F. Berlin
235 Arps Hall
1945 High Street
Columbus, Ohio 43210

APPENDIX C

ADMINISTRATOR'S SUPPORT LETTER

ADMINISTRATOR'S SUPPORT LETTER

I support the PhD project entitled ***Science Teachers' Beliefs***, to be conducted with the science teachers in ______________________________ , Ohio.

Dr. Donna F. Berlin, Principal Investigator, and her Co-Investigator, Tricia Reda Kerr, have provided a written proposal detailing the purpose of the project; a brief description of the project, including duration; a brief description of the subjects; data collection and analysis procedures; a copy of the survey; and a description of the measures to be taken to protect and ensure confidentiality. Possible benefits of the study have been described.

I acknowledge that I have had the opportunity to obtain additional information regarding the study and that any questions I have raised have been answered to my full satisfaction. Furthermore, I understand that I am free to withdraw my support at any time and discontinue participation in this PhD project.

Finally, I acknowledge that I have read and fully understand the support form. I sign it freely and voluntarily. A copy has been given to me.

Date: __________________ Signed: ____
District Administrator

Date: __________________ Signed: ____
OSU Faculty Advisor/Principal Investigator

Date: __________________ Signed: ____
Co-Investigator

APPENDIX D

DESCRIPTION OF RESEARCH

Description of Research

The survey entitled "Science Teachers' Beliefs" is a component of my Ph.D. degree program. The intention of my research is to validate a survey instrument for use in further studies. While several efficacy instruments already exist, none that I investigated referred to curricular recommendations made in the *National Science Education Standards*, especially those standards relating to inquiry-based instruction and learning strategies.

The targeted participants are teachers in the middle grades – defined here as grades 5 through 10 – for whom science is at least 50% of their teaching load. They are asked to respond to a Likert–type scaled survey related to beliefs about their own teaching of science and the relationship between their beliefs about teaching and their beliefs about students' achievement in science. Additional sections seeking demographical information are vital to the data analysis. The survey takes approximately 30 to 40 minutes to complete, and no follow-up is included in the design.

All responses from the teachers will be kept in strictest confidence. No names or any other identifying information will be used in any reports; findings are reported from aggregated data. Teachers are not required to respond to the survey, and participation in data collection is voluntary. Respondents are asked to sign a permission form for explicit permission to use their responses in the study, and the permission forms will be kept until the conclusion of the study

The teachers' responses will be coded by numbers and analyzed using the Statistical Package for Social Sciences. Descriptive statistics including frequencies, means, standard deviations, and variances will be computed. Inferential statistics including correlations and multivariate analyses of variance will be used to analyze relationships. No names or identifying information linked to individuals will be reported. All forms, questionnaires, and corresponding individual data will be destroyed at the end of the study.

Tricia Reda Kerr
Department of Math, Science, and Technology Education
College of Education
The Ohio State University

Kerr.108@osu.edu

APPENDIX E

SURVEY: TEACHER BELIEFS

Science Teachers' Beliefs (Reda Kerr, 2005)

Introduction

Thank you for agreeing to help me with my dissertation research. The following questionnaire will provide information about what teachers believe about their abilities to teach science and how those beliefs are related to student achievement.

A unique characteristic of this survey is that it is based on national science education standards. Please complete the survey from your perspective as a science teacher, and please provide as much information as possible on the demographics page.

No personally identifying information will be associated with responses; all responses will be coded by number and remain confidential. Dr. Donna Berlin, my research advisor, and I are the only people with access to information connecting an individual's identity with the coded responses. Respondents' personal information will be destroyed at the end of my research project.

Again, thank you for your help.

Tricia Reda Kerr
The Ohio State University
School of Teaching and Learning
Math, Science, and Technology Education Program

Part I: School Demographics

How would you characterize your school? **Please check all that apply.**

☐ Public	☐ Urban	☐ Small (<400 students)	☐ Low % on financial assistance (<25% of families)
☐ Private	☐ Suburban	☐ Medium (400-1000 students)	☐ Medium % on financial assistance (26% - 60% of families)
☐ Other	☐ Rural	☐ Large (>1000 students)	☐ High % on financial assistance (>60% of families)

Please continue on the following page.

Part II: Professional Information

Name ______________________ **Year of Birth** __________ **Today's Date:** ______________ **Gender:** ☐ **Female** ☐ **Male**

Most advanced degree completed ______________ **Number of years as a classroom teacher** ______________

Year degree completed ______________ **Number of years teaching grades 5-10 science** ______________

Professional preparation <u>**PLEASE CHECK ALL THAT APPLY**</u>

Major/Area(s) of Concentration: ☐ Mathematics ☐ Reading/Literacy ☐ Science ☐ Social Studies ☐ Other ____________

Grade Level of Initial Certification/ Licensure ☐ PreK-3 ☐ K-8 ☐ 1-8 ☐ 4-9 ☐ 7-12 ☐ 9-12 ☐ Other ____________

Preservice Teacher Education Program
(That resulted in your initial certification/licensure) ☐ **Undergraduate program (e.g., BS or BA)** ☐ **Graduate program (e.g., MEd)**

Approximate total number of graduate hours completed <u>after</u> most advanced degree ☐ 0 ☐ 1-5 ☐ 6-10 ☐ 11-15 ☐ More than 15 Were they: ☐ Quarters ☐ Semesters ☐ Mixed

Science subjects taught during your career ☐ Biology ☐ Chemistry ☐ Physics ☐ Physical Science ☐ Earth Science ☐ Integrated Science ☐ General Science ☐ Other ______________

Professional development experiences: <u>After initial certification/licensure</u> <u>**PLEASE CHECK ALL THAT APPLY.**</u>

Graduate level college or university	**Science ***		**Science Education * ***	
Classes or courses:	☐ Presenter	☐ Participant	☐ Presenter	☐ Participant
Online classes or continuing education units:	☐ Presenter	☐ Participant	☐ Presenter	☐ Participant
Conferences	**Science ***		**Science Education * ***	
Local/District/State:	☐ Presenter	☐ Participant	☐ Presenter	☐ Participant
National:	☐ Presenter	☐ Participant	☐ Presenter	☐ Participant
International:	☐ Presenter	☐ Participant	☐ Presenter	☐ Participant

Did you attend or present at a **science *** conference (local, district, state, national, or international level) **before** you became certified or licensed for the first time? ☐ Yes ☐ No

Did you attend or present at a **science education * *** conference (local, district, state, national, or international level) **before** you became certified or licensed for the first time? ☐ Yes ☐ No

* "Science" refers to discipline-specific content such as geology, chemistry, or anatomy; and conferences such as those sponsored by the American Chemical Society.

* * "Science Education" refers to pedagogy and discipline-specific information in the content area, such as, science methods classes, or conferences such as those sponsored by the Science Education Council of Ohio, the National Science Teachers Association, or National Association of Biology Teachers.

Please continue on the following page.

Part III: Preservice Science Teacher Education Programs

Directions: The following list relates to learning how to teach science. **Please indicate, by circling your choice**, how important you think it is to address these in a preservice science teacher education program. **Note:** The items are arranged in alphabetical order to avoid the appearance of ranking.

	Not Important		Neutral		Very Important
1. Alternative assessments in science (e.g., creating graphics, writing open-response, or designing projects)	1	2	3	4	5
2. Classroom management for interactive science	1	2	3	4	5
3. Cooperative learning groups: Appropriate use in science	1	2	3	4	5
4. Diverse learners and special populations in the science learning environment	1	2	3	4	5
5. Hands-on activities: Designing and/or modifying	1	2	3	4	5
6. Inquiry-based science: Using it to its best advantages	1	2	3	4	5
7. Lesson planning for inquiry-based science	1	2	3	4	5
8. Motivation techniques for reluctant learners of science	1	2	3	4	5
9. Parental involvement: Extending science beyond the classroom	1	2	3	4	5
10. Prior knowledge and experiences related to science: Identifying what students already know	1	2	3	4	5
11. Safety in the science classroom, lab, and in the field	1	2	3	4	5
12. Science concepts: Identifying and teaching concepts appropriate for students	1	2	3	4	5
13. Science education textbooks: Evaluation and selection	1	2	3	4	5
14. Science equipment and materials: Evaluation and selection	1	2	3	4	5
15. Science experiments and investigations: Designing and/or identifying	1	2	3	4	5
16. Science teaching strategies: Identifying strategies and developing one's skills	1	2	3	4	5
17. Scientific processes: Their role and their importance in science	1	2	3	4	5
18. Scientific terminology: Its role and its importance in science	1	2	3	4	5
19. Self-evaluation and reflection: Insights and revelations	1	2	3	4	5
20. Teacher accountability and ethics	1	2	3	4	5
21. Technology: Its role and its importance in science	1	2	3	4	5
22. Traditional tests (e.g., multiple choice, true/false, and matching)	1	2	3	4	5

Please continue on the following page.

Part IV: Teaching and Student Achievement

Directions: Please indicate, by circling your choice, your level of agreement or disagreement with the following statements.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. For me, managing the classroom during a hands-on science activity is overwhelming.	SD	D	N	A	SA
2. I feel capable of teaching science to all students.	SD	D	N	A	SA
3. I find it difficult to explain to students why science experiments and investigations come out the way they do.	SD	D	N	A	SA
4. For me, issues of safety in the science classroom are not intimidating.	SD	D	N	A	SA
5. I welcome students' science questions.	SD	D	N	A	SA
6. If I spend more time planning science lessons, students' achievement will improve.	SD	D	N	A	SA
7. I understand science processes well enough to be effective in teaching science.	SD	D	N	A	SA
8. Students' achievement in my science class is a direct result of my effectiveness in teaching science.	SD	D	N	A	SA
9. In my class, students' understanding of science concepts can be measured by how well they perform on traditional tests (e.g., multiple choice, true/false, and matching.)	SD	D	N	A	SA
10. I believe I could teach a class that included advanced science concepts and processes.	SD	D	N	A	SA
11. For me, the lecture is the most effective way to teach science concepts.	SD	D	N	A	SA
12. High achievement of students in my science class depends on memorizing definitions of scientific terminology.	SD	D	N	A	SA
13. I am comfortable teaching inquiry-based science lessons.	SD	D	N	A	SA
14. Generally, I feel secure about teaching science.	SD	D	N	A	SA
15. I don't know how to inspire students to do science.	SD	D	N	A	SA
16. Every student in my class should have a science textbook to ensure high achievement.	SD	D	N	A	SA
17. When the science grades of students improve, it is often due to my having found a more effective teaching approach.	SD	D	N	A	SA
18. I am not very effective in developing science experiments and investigations for students.	SD	D	N	A	SA
19. When a student has difficulty understanding a science concept, I am usually at a loss as to how to help the student understand it better.	SD	D	N	A	SA
20. I know how to teach science concepts effectively.	SD	D	N	A	SA
21. Even if I try very hard, I do not teach science very well.	SD	D	N	A	SA
22. I am effective in guiding students to design their own experiments and investigations.	SD	D	N	A	SA

Please continue on the following page.

Please circle your choice.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
23. I can overcome the inadequacies of students' science background by good teaching.	SD	D	N	A	SA
24. I can involve students in inquiry-based science with teacher-led demonstrations.	SD	D	N	A	SA
25. Typically, I am able to answer students' science questions.	SD	D	N	A	SA
26. Even if I increase my effort in teaching science, some students will show little change in achievement.	SD	D	N	A	SA
27. I understand science concepts well enough to be effective in teaching science.	SD	D	N	A	SA
28. In my class, students who participate regularly in inquiry-based science experiments and investigations have higher achievement scores.	SD	D	N	A	SA
29. When students are underachieving in science in my class, it can be due to my ineffective science teaching.	SD	D	N	A	SA
30. When parents/guardians of my science students comment that their child is showing more interest in science at school, it is probably due to my performance as a science teacher.	SD	D	N	A	SA
31. It makes me uncomfortable to deal with the idea of teaching science to all students.	SD	D	N	A	SA
32. When I plan and use connected, sequenced lessons, my students' achievement in science improves.	SD	D	N	A	SA
33. Students' achievement will be poor when my science lessons do not include hands-on activities.	SD	D	N	A	SA
34. In my class, traditional tests (e.g., multiple choice, true/false, and matching) are the best way to measure students' achievement in science.	SD	D	N	A	SA
35. I feel at ease when students are working in cooperative learning groups during science experiments and investigations.	SD	D	N	A	SA
36. In my class, students' achievement in science will improve when they work in cooperative learning groups.	SD	D	N	A	SA
37. Having adequate science equipment and materials directly results in higher achievement by students in my class.	SD	D	N	A	SA
38. When my low-achieving students progress in science, it is usually due to extra attention I give as their science teacher.	SD	D	N	A	SA
39. Given a choice, I will not invite the principal or supervisor to evaluate my science teaching.	SD	D	N	A	SA
40. I am not confident of my ability to teach science effectively.	SD	D	N	A	SA
41. I am comfortable using a variety of technologies in my science teaching.	SD	D	N	A	SA
42. I wonder if I have the necessary skills to teach science.	SD	D	N	A	SA
43. I will continually find ways to teach science more effectively.	SD	D	N	A	SA
44. When I use technology in my science class, it ensures that students will have higher achievement.	SD	D	N	A	SA

Please continue on the following page.

Please circle your choice.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
45. In my class, students' achievement improves when science instruction is inquiry-based.	SD	D	N	A	SA
46. I am not comfortable using alternative assessments (e.g., creating graphics, writing open-response, or designing projects) in science.	SD	D	N	A	SA
47. Students' achievement improves when I use self-evaluation and reflection about my science teaching.	SD	D	N	A	SA
48. My students' achievement will not improve unless the science class is highly structured.	SD	D	N	A	SA
49. All students can achieve at high levels in my science class.	SD	D	N	A	SA
50. Teaching with hands-on activities in my science class makes me uncomfortable.	SD	D	N	A	SA
51. Low-achieving students can be motivated to improve in my science class.	SD	D	N	A	SA
52. I am comfortable with parental involvement in my science class.	SD	D	N	A	SA
53. My own background in science helps me teach science more effectively.	SD	D	N	A	SA
54. Students who follow safety requirements for science experiments and investigations achieve at higher levels in my class.	SD	D	N	A	SA
55. In my class, learning concepts in depth is directly related to higher student achievement in science.	SD	D	N	A	SA
56. I can teach more effectively if I rely on only the science textbook for lesson plans.	SD	D	N	A	SA
57. I cannot be an effective science teacher without adequate science equipment and materials.	SD	D	N	A	SA
58. In my class, students must learn scientific processes in order to improve their understanding of science concepts.	SD	D	N	A	SA
59. I am comfortable using scientific terminology in my science teaching.	SD	D	N	A	SA
60. I can measure students' depth of understanding of science concepts by using alternative assessments (e.g., creating graphics, writing open-ended responses, or designing projects).	SD	D	N	A	SA

Thank You!