

## The Development and Validation of an Instrument to Measure Preservice Teachers' Self-Efficacy in Regard to The Teaching of Science as Inquiry

**Lori Dira Smolleck**

Department of Education, Bucknell University, Lewisburg, PA, 17837, U.S.A.

**Carla Zembal-Saul**

College of Education, The Pennsylvania State University, State College, PA, 16802, U.S.A.

**Edgar P. Yoder**

College of Agricultural Sciences, The Pennsylvania State University, PA, 16802, U.S.A.

Published online: 17 June 2006

*The purpose of this study was to develop, validate, and establish the reliability of an instrument that measures preservice teachers' self-efficacy in regard to the teaching of science as inquiry. The instrument, Teaching Science as Inquiry (TSI), is based upon the work of Bandura (1977, 1981, 1982, 1986, 1989, 1995, 1997), Riggs (1988), and Enochs and Riggs (1990). Self-efficacy in regard to the teaching of science as inquiry was measured through the use of a 69-item Likert-type scale instrument designed by the author of the study. Based on the standardized development processes used and the associated evidence, the TSI appears to be a content and construct valid instrument with high internal reliability for use with preservice elementary teachers to assess self-efficacy beliefs in regard to the teaching of science as inquiry.*

### Introduction

Reform of science education in the United States is a difficult, multifaceted task. Within the realm of science education reform, emphasis relies heavily on the importance of scientific inquiry experiences for K-12 learners. *The National Science Education Standards*, (NSES; [National Research Council \[NRC\], 1996](#)), a current reform document aimed at improving scientific literacy for all, is striving to achieve this challenging goal by emphasizing an approach to teaching and learning about science and highlights scientific inquiry as a prominent feature. The National Committee on Science Education Standards and Assessment (NRC) has asserted that students should “engage in aspects of inquiry as they learn the scientific way of knowing the natural world, but they should also develop the capacity to conduct complete inquiries” (p. 23). The problem, however, exists in the fact that many teachers report that they have never experienced teaching or learning science as inquiry (Kleine et al., [2002](#); Windschitl, [2002](#)). This lack of experience is particularly true at the elementary level.

If science reform is going to be successful and our elementary children are to be provided with effective science instruction, preservice teachers must first be provided with opportunities to experience success as learners of science in reform-oriented contexts (Enochs & Riggs, 1990; Kleine et al., 2002; Riggs, 1988). They themselves must experience first-hand how learning science as inquiry takes place within an elementary school setting (Windschitl, 2002). Based on the idea of Bandura's (1977) social learning theory, if preservice teachers can experience success within a science methods course, then they are more likely to model effective instruction within their own elementary classroom, which, in turn, may promote the success of their elementary students in the area of science. Thus, teacher preparation experiences have become a logical target for change.

While teachers should utilize an arsenal of different strategies when teaching science, the NSES (NRC, 2000) described five essential features of classroom inquiry that apply across all grade levels.

1. Learner engages in scientifically oriented questions,
2. Learner gives priority to evidence in responding to questions,
3. Learner formulates explanations from evidence,
4. Learner connects explanations to scientific knowledge, and
5. Learner communicates and justifies explanations. (p. 29)

These essential features introduce important aspects of science to students while simultaneously assisting them in developing knowledge in regard to specific science concepts. Within each of the five essential features of classroom inquiry, there are variations that are labeled within a continuum of inquiry experiences. To determine whether an experience is categorized as full or partial inquiry, one must consider the amount of student and teacher involvement. Partial inquiries have greater teacher involvement and less student involvement, whereas full inquiries have greater student involvement, and less teacher involvement. For example, when students are engaging in scientifically oriented questions, the amount of inquiry will vary depending on the origin of the question. If the question was presented by the teacher, the amount of inquiry will be less than if the students developed their own questions to investigate. In essence, within any classroom, a lesson may be more or less student directed, depending on the variation of the features implemented.

Given the nature of teaching and learning science as inquiry, it is important to point out that this type of learning and teaching is not a neat and tidy process. The NRC (2000) described a more specific set of variations to encompass inquiry learning. The NRC proposed a definition that is derived from ". . . the abilities of inquiry, emphasizing questions, evidence, and explanations within a learning context" (p. 24). At the center of this definition are the five essential features of classroom inquiry.

These features draw attention to students engaging in scientifically oriented questions and giving priority to evidence when formulating explanations. In addition, students evaluate their explanations in light of alternative explanations and then communicate their proposed explanations to others (NRC, 2000).

Although, scientific inquiry can be defined and described in a number of ways, for the purposes of this research, we will be using the definitions set forth by the NRC (2000). Our definition draws upon the essential features of classroom inquiry as the essence of inquiry, particularly the notion of giving priority to evidence and explanation.

### **Theoretical Underpinnings**

In this section, we discuss how previously established research provided a foundation for this project. These areas include the importance of scientific inquiry experiences, social learning theory, self-efficacy, and self-efficacy within the context of science education. The focus of this research on beliefs is a result of the relationship between attitude and subsequent behavior (Bandura, 1986). Bandura's theory of social learning describes two dimensions of efficacy beliefs—personal self-efficacy and outcome expectancy—upon which behavior is based. As explained by social learning theory, if teachers do not have successful experiences teaching or learning science as inquiry, it is unlikely that these teachers will implement science as inquiry in their elementary science classrooms (Bandura, 1977).

### **Factors Driving Science Education Reform**

Given the current trends in contemporary science education reform, science as inquiry is considered to be an important part of the restructuring. A rich learning environment, with a focus on inquiry-based learning, creates opportunities for children to internalize or transform new information, which then allows students to "...create and expand their individual cognitive structures" (Lee & Krapfl, 2002, p. 250). This type of learning supports conceptual understanding as opposed to rote learning of science concepts and facts (Kleine et al., 2002; Lee & Krapfl, 2002).

Children should know how to pursue their own questions about the world around them. This pursuit, however, does not happen naturally in the classroom, and students will need to be supported in their attempts to understand phenomena. When science is taught through the process of inquiry, children will have the opportunity to pose questions and seek answers based on observation and exploration. Students can then use the evidence gathered throughout this process to answer their own questions, as well as future questions that may arise. Inquiry allows students the opportunity to explore, yet simultaneously requires them to learn something about how science is done (Drayton & Falk, 2001; NRC, 2000).

This inquiry approach to teaching and learning science allows the teacher to become more facilitative in their instruction and the students to become more self-directed. As a result of this shift from a more teacher-centered classroom to a more student-centered classroom, students are able to establish "long-term conceptual understandings of science" (Kleine et al., 2002, p. 39). There is substantial empirical and theoretical evidence to support the assertion that inquiry-based science instruction is a starting point for personal construction of meaning and can also lead to higher achievement for students (Anderson, 1997; Freedman, 1997; Von Secker

& Lissitz, 1999). As noted in theoretical claims supported by empirical evidence, "...greater emphasis on inquiry-based teaching is associated with higher science achievement" (Von Secker, 2002, p. 156). Scientific inquiry introduces students to the content of science, as well as the processes of investigation. "It provides the logical framework that enables students to understand scientific innovation" (Drayton & Falk, 2001, p. 25).

Although the push for science as inquiry within elementary classrooms is enormous, unfortunately it has yet to become a consistent feature of science classroom practice (Damjanovic, 1999; Drayton & Falk, 2001; Weiss, Banilower, McMahon, & Smith, 2001; Wells, 1995; Windschitl, 2002). One possible reason for the lack of inquiry teaching in the elementary science classroom could be a reflection of the mismatch between teacher beliefs and the context of science (Windschitl, 2002). The problem with the current teaching of science is that it fails to reflect the changes in science that have occurred over the years; the process of science and the teaching of science have drifted apart. Despite the fact that many classrooms today give the illusion that inquiry is provided, such lessons typically involve structured procedures and the results provided through the textbook, as well as through the classroom experience (Drayton & Falk, 2001; Schwab, 1962). The reasoning behind this lack of visibility lies heavily on the conceptions of teachers (Windschitl, 2002). Many teachers believe that teaching science as inquiry is very difficult and cumbersome to implement and manage within classroom practice. Teachers believe that implementation is obstructed due to such constraints as time and money. They also feel that teaching science as inquiry "is possible only with above-average students" and, therefore, do not attempt to integrate inquiry into their regular education classrooms (Lee & Kraphfl, 2002; as cited in Windschitl, 2002, p. 115).

Although some of these reasons for the lack of science as inquiry experiences for elementary children may be viable, teachers themselves need to feel confident utilizing inquiry, both as learners and as teachers, so students can learn to participate in the processes of science (Kleine et al., 2002; Windschitl, 2002). One vehicle for achieving this level of confidence is through the investigation of how self-efficacy may impact teacher practice.

### **Social Learning Theory and Self-Efficacy Defined**

One way to explain why some teachers choose to eliminate science instruction from their daily routine is through social learning theory. Bandura's social learning theory suggests that "people develop a generalized expectancy concerning action-outcome contingencies based upon life experiences" (Riggs, 1998, p. 2). "The strength of peoples' convictions in their own effectiveness is likely to affect whether they will attempt to cope with given situations; hence, perceived self-efficacy influences choice of behavior settings" (Bandura, 1977, p. 193). When individuals "...judge themselves as capable of handling situations that would otherwise be intimidating, they become involved in activities and behave assuredly, however when situations exceed one's own coping skills, individuals tend to fear and avoid" these difficult situations (Bandura, p. 194).

It has been found that efficacy expectations are presumed to influence levels of performance. Research also has indicated the predictive power of one's sense of self-efficacy on subsequent performance: "Efficacy expectations are a major determinant of people's choice of activities, how much effort they will expend, and of how long they will sustain effort in dealing with stressful situations" (Bandura, 1977, p. 194).

Developing self-confidence as a teacher of science is crucial, especially for preservice elementary teachers. In a national survey, Weiss et al. (2001) indicated that elementary teachers teach science an average of 25 minutes per day as opposed to about 114 min per day for reading and language arts. Several reasons are given to explain this phenomenon, including "lack of a strong background in science content, inadequate facilities and equipment, the congested curriculum, poor instructional leadership, and teacher attitude" (Enochs & Riggs, 1990, p. 694). Although elementary teachers are responsible for teaching all content areas, it is often noted that many elementary teachers do not feel comfortable teaching science (Martin, 2000). The difference between teachers that allow for more science instruction within their elementary classrooms and those who do not may be related to their self-confidence. Those teachers who do not believe in their ability to teach science (low self-efficacy) may avoid science instruction whenever possible (Enochs, Scharmann, & Riggs, 1995). Because of the relationship between beliefs, attitudes, and behavior with regard to elementary science teaching, this notion supports the conclusion that efficacy beliefs are potentially powerful variables that can influence the amount of science instruction time, as well as the achievement of students in science, at the elementary level.

Because Bandura (1981) described self-efficacy as a situation-specific construct, we chose to create an instrument that measures self-efficacy in regard to the teaching of science as inquiry. Although there are many instruments used to measure teacher self-efficacy in general terms, specificity was necessary within this study. Teacher efficacy beliefs appear to be dependent upon the specific teaching situations (Riggs & Enoch, 1990). Because elementary teachers typically teach all subjects—and may not feel equally effective at teaching all of them—a subject-specific instrument was more informative for the purposes of this study.

As noted earlier, Bandura (1977) asserted that the most complete prediction of human behavior can be derived from knowledge of both self-efficacy and outcome expectancy variables. Personal self-efficacy is "a judgment of one's ability to organize and execute given types of performances, whereas an outcome expectation is a judgment of the likely consequence such performances will produce" (p. 21). These principles underlie the development of the instrument in this study. The instrument we created was based on contemporary ideas about inquiry, as well as grounded in the fundamental ideas of Bandura, particularly the notion of self-efficacy being a context-specific construct.

The level of motivation an individual has for a given situation, their associated feelings toward the situation, and their subsequent behaviors are "based more on what they believe, rather than on what is objectively true" (Bandura, 1997, p. 2). "Unless people believe they can produce desired effects by their actions, they have

little incentive to act” (Bandura, 1997, p. 3). Bandura (1995) explained the varying effects of perceived sense of efficacy as follows:

*A strong sense of efficacy enhances human accomplishment and personal well-being in many ways. People with high assurance in their capabilities in given domains approach difficult tasks as challenges to be mastered rather than as threats to be avoided. Such an efficacious outlook fosters intrinsic interest and deep engrossment in activities. These people set themselves challenging goals and maintain strong commitment to them. They heighten and sustain their efforts in the face of difficulties. They quickly recover their sense of efficacy after failures or setbacks. They attribute failure to insufficient effort or to deficient knowledge and skills that are acquirable. They approach threatening situations with assurance that they can exercise control over them. Such an efficacious outlook produces personal accomplishments, reduces stress, and lowers vulnerability to depression. (1995, p. 11)*

On the other hand,

*People who have a low sense of efficacy in given domains shy away from difficult tasks, which they view as personal threats. They have low aspirations and weak commitment to the goals they choose to pursue. When faced with difficult tasks, they dwell on their personal deficiencies, the obstacles they will encounter, and all kinds of adverse outcomes rather than concentrate on how to perform successfully. They slacken their efforts and give up quickly in the face of difficulties. They are slow to recover their sense of efficacy following failure or setbacks. Because they view insufficient performance as deficient aptitude, it does not require much failure for them to lose faith in their capabilities. They fall easy victim to stress and depression. (1995, p. 11)*

These quotes from Bandura support the factors previously discussed that contribute to the omission of science in the elementary classroom. Understanding the foundation upon which these quotes were built will help educators to more thoroughly understand why teachers choose to eliminate science from their curriculum and how these obstacles can be overcome.

### **Self-Efficacy Within the Context of Science Education**

To more fully address the idea that teacher efficacy is context specific, Aston, Buhr, and Crocker (1984, as cited in Tschannen-Moran & Woolfolk-Hoy, 2001) developed a series of vignettes. These vignettes described situations a teacher would be likely to encounter and required the teacher to make judgments regarding their effectiveness in responding to the situations (Tschannen-Moran & Woolfolk-Hoy, 2001). A sample item follows:

*A small group of students is constantly whispering, passing notes and ignoring class activities. Their academic performance on tests and homework is adequate and sometimes even good. Their classroom performance, however, is irritating and disruptive. How effective would you be in eliminating their disruptive behavior? (p. 788)*

Although this instrument was useful in that it addressed the assumption that teacher efficacy is context specific, it did not receive wide acceptance and has not been used in any other research studies.

Drawing upon this work, Gibson and Dembo (1984) developed and validated the Teacher Efficacy Scale (TES), a questionnaire of 30 items using a 6-point Likert-type scale. This instrument was designed to measure teachers' self-efficacy beliefs by addressing the areas of their effort, skill, training, and experience. While the TES was believed to measure both general teaching efficacy (GTE) and personal teaching efficacy (PTE), other research did not substantiate this distinction. The TES was criticized for not clearly capturing the dimension of personal efficacy as described by Bandura's definition of the self-efficacy construct. In addition, the TES instrument developed by Gibson and Dembo was a global measure of the two efficacy factors. This caused concern because it was not consistent with Bandura's conception of efficacy as a situation-specific construct. This concern later encouraged others to further develop situation-specific self-efficacy instruments.

To add to the literature base on self-efficacy with attention to Bandura's description of a situation-specific construct, Riggs (1988) extended the work of Gibson and Dembo (1984). Riggs developed and validated an instrument to measure teachers' personal self-efficacy and outcome expectancy beliefs for science teaching and learning. This instrument was entitled the Science Teaching Efficacy Beliefs Instrument (STEBI).

This work by Riggs was further extended by Enochs and Riggs (1990) to address preservice elementary science teachers' self-efficacy. This instrument was entitled Science Teaching Efficacy Belief Instrument for Prospective Teachers (STEBI-B). To accomplish this task, the Riggs (1988) "STEBI-A was modified from an inservice orientation to that of preservice" (Enochs & Riggs, p. 696). Items were reworded in the future tense to allow "...for the construct to be viewed in a different situational context" (Enochs & Riggs, p. 696). Both the STEBI-A and the STEBI-B have become widely used in science education to inform teacher educators about the science beliefs of prospective teachers.

Although the Science Teaching Efficacy Belief Instrument for Prospective Teachers (STEBI-B) has been determined to be a valid and reliable instrument used to investigate preservice elementary science teachers' self-efficacy, it does not measure teaching efficacy in teaching science as inquiry. For example, the following statements, "I understand science concepts well enough to be effective in teaching elementary science" and "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" (Riggs & Enochs, 1990, p. 635) seem like effective statements for measuring science teaching; however, they do not capture the essence of scientific inquiry. Although

this measure was a reflection of how science was taught during its time period, it does not capture the essential features of classroom inquiry. This measure was developed before reform documents were published; therefore, it is not a reflection of contemporary education reform.

In addition, the statements, “I find it difficult to explain to students why science experiments work,” and “I am typically able to answer students’ science questions,” again do not support the process of scientific inquiry (Riggs & Enochs, 1990, p. 635). These statements infer that the teacher must explain to students the many phenomena that exist in the world of science and that every experiment has a certain procedure that must be followed to arrive at the correct answer. These statements do not draw attention to students engaging in scientifically oriented questions and giving priority to evidence when formulating explanations. Furthermore, the statements do not involve students in evaluating their explanations in light of alternative explanations and then communicating their proposed explanations to others. Although the STEBI was a useful tool in its time, the current standards associated with contemporary science education reform require a new instrument to fit the ever-changing complexities of science education.

More recently, Tschannen-Moran and Woolfolk-Hoy (2001) attempted to begin work on a new measure of efficacy. This new measure, named the Ohio State Teacher Efficacy Scale (OSTES), is said to be “. . .superior to previous measures of teacher efficacy” (p. 801). This statement was based on the assertion that the measure captures a broad range of capabilities necessary for good teaching. However, it is not so specific that it renders itself useless for comparison of teachers in different domains or contexts. The OSTES is considered to be reasonably valid and reliable. The researchers have asserted that this measure “. . .moves beyond previous measures to capture a wide range of teaching tasks” (p. 801). In addition, this instrument is unique in that it addresses a broad range of teaching responsibilities, including assessment, meeting individual student needs, motivating student engagement and interest, and addressing student misconceptions.

### **Significance of the Study**

Although various researchers have set out to improve teacher education in admirable ways, few of these attempts have involved the relationship between self-efficacy beliefs and practice. Recalling Bandura’s (1977, 1986, 1995) assertions that self-efficacy beliefs have a powerful influence over one’s behavior, it is important for teacher educators to further investigate the influence of teacher beliefs on classroom practice. As a result of the relationship among beliefs, attitudes and behavior, the purpose of this study was to create an instrument that measures preservice teachers’ self-efficacy in regard to the teaching of science as inquiry. Additionally, given the current trend of contemporary science education reform, there is not an instrument to measure self-efficacy and its impact on the teaching of science as inquiry. As a result of the current trends in science education—and the renewed interest in inquiry—there is a need to focus on the teaching of science as inquiry. Due to the shift in science education, it was our goal to contribute to the science education



literature by addressing the ideas of where self-efficacy and inquiry science teaching connect.

As such, by developing, validating, and establishing the reliability of an instrument to measure self-efficacy beliefs of preservice elementary teachers with regard to the teaching of science as inquiry, the current reform efforts will also be addressed. We hope this instrument will provide a foundation through which researchers can identify certain individuals and investigate the connections between beliefs and actual teaching behaviors and classroom practices. Through the completion of more extensive research coupled with this instrument, science educators may come to more clearly understand the connection between teacher beliefs and the teaching of elementary science as inquiry.

## Method

### Participants

Data were collected from elementary education majors during their senior year at a central Pennsylvania university. Participants were enrolled in a science methods course during the time of data collection. The Teaching Science as Inquiry (TSI) instrument was administered to 190 prospective elementary school teachers in six sections of a science methods course during the week of September 8, 2003, and then again during the week of December 1, 2003, within a university classroom setting. This group of participants represents the intended target population for the final instrument. The intended population would, therefore, include preservice elementary science teachers and beginning practicing science teachers.

### Development of the TSI

A 13-step process was utilized to complete this study. This 13-step plan described below was used to develop and build validity and high reliability into the TSI.

*Step 1: Defining the Construct.* According to Graziano and Raulin (2000), a Construct “is an idea constructed by the researcher to explain events observed in a particular situation. They are explanatory fictions because, in most cases, we do not know the real reason for a particular event. Once formulated, constructs are used as if they are true to predict relationships between variables in situations that had not previously been observed” (p. 419).

In this study, the instrument to be developed measured preservice teachers’ self-efficacy in regard to the teaching of science as inquiry. Specifically, the two dimensions of self-efficacy to be measured are personal self-efficacy and outcome expectancy as defined by Bandura (1977).

Teaching and learning science as inquiry as recognized by the National Science Education Standards (NRC, 2000) involves five essential features:

1. Learner engages in scientifically oriented questions,
2. Learner gives priority to evidence in responding to questions,
3. Learner formulates explanations from evidence,
4. Learner connects explanations to scientific knowledge, and
5. Learner communicates and justifies explanations. (p. 29)

*Step 2: Item Preparation, Version 1.* The first phase involved the development of a preliminary instrument to assess preservice teachers self-efficacy in regard to inquiry science teaching. To develop this preliminary version of the instrument, the researchers utilized the text, *Inquiry and the National Science Education Standards* (NRC, 2000). To devise items representative of the construct to be studied, the researchers first developed Version 1 of the TSI. This version of the instrument was composed of a broad set of 31 statements to capture the nature of teaching science as inquiry. Example statements included “allowing student interest to guide the curriculum” and “provide opportunities for students to discuss the experiments in which they participated.” Although the items represented in Version 1 of the instrument followed the standards set forth by the NSES (NRC), they were only a broad set of items that clearly needed to be more thoroughly developed.

*Step 3: Content Validity, Version 1.* A panel composed of six faculty members from the University of Florida, The Pennsylvania State University, and the University of Missouri and three graduate students from The Pennsylvania State University representing the areas of science education and self-efficacy research was assembled for the purpose of judging the content validity of Version 1 of the instrument. Each member of the panel had elementary school, middle school, or both science classroom experience. Independent preliminary feedback from each member of the panel was collected and used as a means to revise the items. Overall, the two following broad themes emerged: basic grammatical revisions and content revisions. To address these themes, many of the 31 statements were revised to become more specific and more clear. The most significant revision made to Version 1 was that several statements were added to the instrument. The goal in adding these items was to more thoroughly represent the ideal of teaching science as inquiry. The panel also suggested including items that represent the five essential features of classroom inquiry. As a result of the activities identified in the first three steps, items from Version 1 were revised. This provided a basis for final item preparation for the next phase of the development process.

*Step 4: Content Validity, Version 2.* Upon completion of Version 2 of the instrument, a letter that explained the review process and the 81 draft items of Version 2 were submitted to the panel of nine experts. Directions, as well as a definition of “inquiry,” were provided with the instrument to ensure that all reviewers responded to the instrument in the same manner. The panel reviewed each of the items independently for clarity and comprehension. They were also asked to

categorize each of the 81 items into the following five groups, which pertain to the five essential features of classroom inquiry (listed in the “Introduction” and “Method” sections).

In addition to categorizing the 81 statements, panel members were also invited to provide suggestions for rewording or rephrasing the statements. Additionally, the panel was also encouraged to identify statements they would like to see added to the list. Example statements included “I will be able to engage students in designing the learning environment in an attempt to allow for diversity of problems and methods,” “I possess the ability to ‘let go’ and allow students to devise their own problems to investigate,” and “The majority of evidence is derived from instructional materials such as a text book.”

Comments for improving the items and the categorizations were recorded directly on the instrument. Feedback was collected by the researcher and used to revise the items. Information concerning the categorization of each of the items was used to verify if the items represented the intended essential features of classroom inquiry. The researcher analyzed the panel’s feedback to identify patterns among the data. Items with vast agreement remained on the instrument, whereas items with vast disagreement were revised according to the panel’s suggestions or removed from the instrument if suggestions were not provided. Revisions also included basic grammatical corrections and content clarification. In particular, the reviewers suggested revising some of the statements to better represent the five essential features of classroom inquiry previously listed.

These revised items were then submitted to the panel of faculty members and graduate students for further review. The faculty members and graduate students revised until they judged that the clarity and comprehension was achieved for each of the items. After two rounds of review by the faculty members and graduate students, 94 items were consequently identified for Version 3.

As a result of the activities identified in Step 4, items from Version 2 were revised. This provided a basis for final item preparation for the next phase of the development process.

*Step 5: Content Validity, Version 3.* Using the newly revised version of the instrument, another round of construct validity was conducted. The reviewers for this round consisted of three faculty members from The Pennsylvania State University representing science education and self-efficacy research. These faculty members were given each of the items on the instrument one at a time and asked to place each of the items on a larger representation of the essential features of classroom inquiry and their variations (NRC, 2000, p. 29). Throughout this process, the researcher asked the reviewers to verbally convey their thought processes and reasoning behind each of their placements. During this time, the members of the review panel collectively provided feedback to the researcher. To ensure that all feedback would be considered and reflected in the next version of the instrument, the researcher audiotaped and then transcribed this meeting.

*Step 6: Revision of Items.* As a result of the process discussed in Step 5, Version 3 of the instrument was revised. Most of the revisions consisted of rephrasing the items to better represent the construct of science as inquiry, as well as Bandura's (1977) definition of self-efficacy. Specifically, upon suggestions of the faculty members, the researcher rephrased many of the items to capture the two dimensions of self-efficacy described by Bandura: personal self-efficacy and outcome expectancy. The faculty members also suggested adding seven new items to the instrument in an attempt to better address the definition of self-efficacy and the definition of teaching science as inquiry. Again, during the process of revision, items with vast agreement remained on the instrument, whereas items with vast disagreement were either revised or removed from the instrument. In addition, many items were revised for clarification and clarity.

The activities identified in Steps 5 and 6 provided a basis for final item preparation for the next version of the instrument, Version 4, which consisted of 65 items. Example statements include "As a teacher of science, I will be able to offer multiple suggestions for creating explanations from data," "As a science teacher, I will provide opportunities for students to become critical decision makers when evaluating the validity of scientific explanations," and "I possess the skills necessary for guiding my students toward explanations that are consistent with experimental and observational evidence."

*Step 7: Content Validity, Version 4.* Similar to all of the other versions, content validity was conducted on Version 4 of the instrument. Four faculty members from The Pennsylvania State University representing science education and self-efficacy research were brought together for the purpose of judging each item's representation of Bandura's (1977) two dimensions of self-efficacy. The reviewers analyzed each of the items independently. Similar to previous content-review processes, comments from each reviewer regarding improving the items were also recorded directly on the instrument. Feedback was collected by the researcher and used to revise the items. As a result of this review process, it was clear that there were more personal self-efficacy items than outcome expectancy items. Each reviewer offered suggestions pertaining to how several of the items could be transformed from personal self-efficacy to outcome expectancy by simply changing the wording of the item. These comments were used as a means to revise and ensure that there was equal distribution of items within the two dimensions of self-efficacy. In addition, a few of the items were revised for clarity and comprehension.

As a result of the activities identified in Step 7, items from Version 4 were revised. This provided a basis for final item preparation for the next phase of the development process—the creation of Version 5.

*Step 8: Content Validity, Version 5.* Version 5 of the instrument also underwent a round of content validity. During this time, faculty members from The

Pennsylvania State University reviewed each of the items to ensure that there was a balance of items within the construct of self-efficacy as defined by Bandura (1977). Specifically, the reviewers viewed each of the items to ensure that there was a relatively equal distribution between the number of items representing personal self-efficacy and the number of items representing outcome expectancy. Each reviewer analyzed the items independently and again made comments directly on the instrument. Basic grammatical and content revisions were made to the items. In addition, revisions were made to enhance the comprehension and clarity of the items.

As a result of the activities identified in Step 8, items from Version 5 were revised. As a result of this round of content validity, revisions were made to Version 5 of the instrument and, consequently, Version 6 was created.

*Step 9: Content Validity, Version 6.* Version 6 of the instrument also underwent a round of content validity. During this time, faculty members from The Pennsylvania State University again reviewed each of the items to ensure that there was a balance of items within the construct of self-efficacy as defined by Bandura (1977). In addition, the reviewers read each of the items to ensure that each clearly illustrated teaching science as inquiry. Specifically, the reviewers again placed each of the items into categories representing the 24 variations of the essential features of classroom inquiry (NRC, 2000, p. 29). Once complete, each of these cells were analyzed to ensure that there was a relatively equal distribution between the number of items representing personal self-efficacy and the number of items representing outcome expectancy within each of the 24 cells. While completing this content validity process, each reviewer analyzed the 65 items independently and again made comments directly on the instrument. Basic grammatical and content revisions were made to the items. In addition, revisions were made to enhance the comprehension and clarity of the items.

As a result of this round of content validity, revisions were made to Version 6 of the instrument and, consequently, Version 7 was created. Table 1 summarizes the results for the item distribution for Version 7 of the instrument. Each cell represents the essential features of classroom inquiry and their variations. The typeface (italics or underline) of each item indicates personal self-efficacy and outcome expectancy.

*Step 10: Administration of Version 7.* Version 7 of the instrument contained 69 items and was administered to the 190 preservice elementary teachers in six sections of a science methods course during the week of September 8, 2003. Of the 190 participants, 91% were female and 9% were male. These groups represented the intended population for the final instrument.

*Step 11: Analysis of Data, Version 7.* The data obtained from administering the 69-item Version 7 to the science methods classes were used to identify the items

**Table 1***Distribution of Items*

	A	B	C	D
Learner engages in scientifically oriented questions	4, <u>19</u> , <u>25</u>	<u>7</u> , <u>11</u> , <u>51</u>	37, 38, <u>48</u> , <u>66</u>	18, <u>21</u> , <u>27</u> , <u>45</u> , <u>46</u>
Learner gives priority to evidence in responding to questions	36, <u>57</u> , <u>58</u>	5, <u>13</u> , <u>17</u> , <u>30</u>	<u>8</u> , <u>44</u> , <u>49</u> , <u>53</u>	<u>29</u> , <u>40</u> , <u>47</u> , <u>52</u> , <u>54</u>
Learner formulates explanations from evidence	2, <u>10</u> , <u>34</u> , <u>35</u> , <u>39</u>	<u>20</u> , <u>26</u> , <u>28</u>	1, <u>31</u> , <u>55</u>	67, <u>69</u>
Learner connects explanations to scientific knowledge	3, <u>15</u> , <u>61</u> , <u>63</u>	14, <u>22</u> , <u>24</u>	23, <u>41</u> , <u>43</u>	XXXXXXXX
Learner communicates and justifies explanations	<u>6</u> , <u>12</u> , <u>33</u>	<u>9</u> , <u>16</u> , <u>32</u> , <u>59</u> , <u>64</u> , <u>65</u>	<u>50</u> , <u>60</u> , <u>62</u>	42, <u>56</u> , <u>68</u>

*Note.* Italics: personal self-efficacy; personal self-efficacy total = 34; underline: outcome expectancy; outcome expectancy total = 35

to be included in the TSI (using SPSS, Version 11.0.4). The following guiding question was developed for this purpose:

*What is the most reliable and valid combination of items to compose the TSI for the purposes of assessing preservice elementary teachers' self-efficacy beliefs in regard to teaching science as inquiry and the two dimensions of self-efficacy: personal self-efficacy and outcome expectancy?*

This question required the researcher to use data from Version 7 to examine the construct validity of the items and the contributions each item made to the reliability of the instrument. Hence, data from the Version 7 items were examined for evidence of construct validity. Item score to total test score correlation and item contribution to total test reliability were used to identify the strongest items and, therefore, eliminate those that were not positively contributing to the overall reliability of the instrument. Item balance across the 24 variations of the essential features of classroom inquiry was also examined to determine the reliability of the instrument. Coefficient alpha, a measure of internal consistency, was utilized to examine the reliability of the instrument. The strongest combination of construct valid and reliable items that had balanced representation within the essential features of classroom inquiry and their variations, were identified using a combination of these procedures.

The ranges on the internal consistency were from .4906 to .7429. Please refer to Tables 2 and 3 for specific internal consistency data. These ranges met or exceeded the requirements set forth by Sax (1974) and Nunnally (1978) pertaining to first generation instrument construction. Outcome expectancy for the category "Learner Connects Explanations to Scientific Knowledge" of the instrument, produced the

**Table 2**

*Reliability Results for Self-Efficacy and the Essential Features of Classroom Inquiry—Version 7*

Essential feature	<i>n</i>	Items	$\alpha$	Item mean	Item variance
Learner engages in scientifically oriented questions	185	7	.6884	3.69	.19
Learner gives priority to evidence in responding to questions	187	8	.6315	3.51	.27
Learner formulates explanations from evidence	185	6	.6189	3.85	.04
Learner connects explanations to scientific knowledge	187	6	.7429	3.89	.04
Learner communicates and justifies explanations	187	7	.7244	4.0	.02

lowest alpha, .4906. Although this alpha is not as high as one would like it to be, omitting any item from this category would only lower the reliability of the instrument. In addition, one factor that contributes to the reliability of a test is the number of items on the test (Anastasi & Urbina, 1997). For this particular analysis, there were only four items in the category resulting in the .4906 alpha. The small amount of items within this category could have possibly accounted for the low reliability.

The data obtained during these analyses were used to arrange items for Version 8 of the TSI. Due to the reliability results as determined by the internal consistency, as well as the correlation data, Version 8 of the TSI consisted of the identical items that were present on Version 7. Revisions made to Version 7 were aesthetic in

**Table 3**

*Reliability Results for Outcome Expectancy and the Essential Features of Classroom Inquiry—Version 7*

Essential feature	<i>n</i>	Items	$\alpha$	Item mean	Item variance
Learner engages in scientifically oriented questions	186	8	.7259	3.50	.12
Learner gives priority to evidence in responding to questions	188	8	.6568	3.51	.46
Learner formulates explanations from evidence	189	7	.6309	3.72	.17
Learner connects explanations to scientific knowledge	189	4	.4906	3.86	.12
Learner communicates and justifies explanations	185	8	.5817	3.56	.18

**Table 4**

*Reliability Results for Self-Efficacy and the Essential Features of Classroom Inquiry—Version 8*

Essential feature	<i>n</i>	Items	$\alpha$	Item mean	Item variance
Learner engages in scientifically oriented questions	184	7	.6579	3.95	.07
Learner gives priority to evidence in responding to questions	181	8	.6583	3.79	.28
Learner formulates explanations from evidence	183	6	.6749	4.05	.03
Learner connects explanations to scientific knowledge	183	6	.7582	4.07	.004
Learner communicates and justifies explanations	183	7	.7566	4.16	.01

nature and were, therefore, done to enhance the readability and visual appearance of the instrument. In addition, those revisions may also contribute to the ease at which participants complete the instrument. Font size was enlarged to increase the participants' ease in reading the items, sentence starters were added to reduce redundancy while reading the items, and shading was added to every other item to provide more ease when completing the instrument.

*Step 12: Administration of Version 8.* A second construct validity and reliability study was conducted on the TSI Version 8 during the week of December 1, 2003. This was done to further develop evidence of the instrument's construct validity and to collect data on the internal reliability and test-retest reliability of the instrument. Version 8 of the instrument contained 69 items and was administered during the week of December 1, 2003, to the 184 preservice elementary teachers in the same six sections of the science methods courses. Of the 184 participants, 90% were female, and 10% were male. These groups again represented the intended population for the final instrument. The resulting data were used in formulating the TSI as described in Step 13, below.

*Step 13: Analysis of Data, Version 8.* Data obtained from the administration of the 69-item Version 8 of the TSI to the science methods classes were used to identify items to be included in the final version of the instrument. Similar to Step 11, again the researcher examined the construct validity of the items and the contributions each item made to the reliability of the instrument. Data from Version 8 of the instrument were examined for evidence of construct validity. Item score to total test score correlation and item contribution to total test reliability were used to identify the strongest items. Item balance across the 24 variations of the essential features of classroom inquiry was also examined to determine the reliability of the instrument. Coefficient alpha, a measure of internal



**Table 5***Reliability Results for Outcome Expectancy and the Essential Features of Classroom Inquiry—Version 8*

Essential feature	<i>n</i>	Items	$\alpha$	Item mean	Item variance
Learner engages in scientifically oriented questions	182	8	.7833	3.75	.15
Learner gives priority to evidence in responding to questions	181	8	.7026	3.73	.29
Learner formulates explanations from evidence	179	7	.6701	3.97	.14
Learner connects explanations to scientific knowledge	183	4	.6034	4.07	.04
Learner communicates and justifies explanations	181	8	.6801	3.81	.19

consistency, was utilized as a means to examine the reliability of the instrument. The strongest combination of construct valid and reliable items that had balanced representation within the essential features of classroom inquiry and their variations, were identified using these procedures in combination. The ranges on the internal consistency were from .6034 to .7833. Please refer to Tables 4 and 5 for specific internal consistency data. These ranges again met or exceeded the requirements set forth by Sax (1974) and Nunnally (1978) pertaining to first-generation instrument construction. In addition, most of these results met the stricter standards described by Anastasi and Urbina (1997) and Isaac and Michael (1997).

One-way analysis of variance (ANOVA) scores on the TSI were compared across the six sections of the science methods course. The .05 level for statistical significance was used to determine if statistically significant differences in the subscale scores existed on the TSI among the six sections. Please refer to Tables 6–9 for further information regarding the ANOVA scores.

Data obtained during these analyses indicated that this pool of items best represented the intended means of the instrument. Due to the reliability results as determined by the internal consistency, as well as the correlation data, the final version of the TSI consisted of the identical items that were present on Version 8. No further revisions were made for the purposes of this research.

## Conclusion

Based on the instrument development processes used and the associated data analysis results, the TSI appears to be a content and construct valid instrument with high to moderate internal reliability and high to moderate test-retest reliability

**Table 6**

*Analysis of Variance Results for Self-Efficacy (SE) and Outcome Expectancy (OE) by Gender—Version 7*

	Sum of squares	<i>df</i>	Mean square	<i>F</i>	Sig
<b>SE1</b>					
Between groups	20.458	1	20.458	2.264	.134
Within groups	1653.820	183	9.037		
Total	1674.278	184			
<b>OE1</b>					
Between groups	2.563	1	2.563	.189	.664
Within groups	2497.587	184	13.574		
Total	2500.150	185			
<b>SE2</b>					
Between groups	19.395	1	19.395	1.877	.172
Within groups	1911.448	185	10.332		
Total	1930.843	186			
<b>OE2</b>					
Between groups	10.551	1	10.551	1.107	.294
Within groups	1773.561	186	9.535		
Total	1784.113	187			
<b>SE3</b>					
Between groups	11.876	1	11.876	2.419	.122
Within groups	898.322	183	4.909		
Total	910.198	184			
<b>OE3</b>					
Between groups	5.474	1	5.474	.764	.383
Within groups	1340.169	187	7.167		
Total	1345.644	188			
<b>SE4</b>					
Between groups	1.989	1	1.989	.347	.557
Within groups	1060.486	185	5.732		
Total	1062.475	186			
<b>OE4</b>					
Between groups	1.609	1	1.609	.629	.429
Within groups	478.556	187	2.559		
Total	480.165	188			
<b>SE5</b>					
Between groups	.081	1	.081	.010	.921
Within groups	1506.610	185	8.144		
Total	1506.691	186			
<b>OE5</b>					
Between groups	39.343	1	39.343	4.209	.042
Within groups	1710.750	183	9.348		
Total	1750.093	184			

**Table 7***Analysis of Variance Results for Self-Efficacy (SE) and Outcome Expectancy (OE) by Gender—Version 8*

	Sum of squares	df	Mean square	F	Sig
<b>SE1</b>					
Between groups	.139	1	.139	.021	.885
Within groups	1199.186	182	6.589		
Total	1199.326	183			
<b>OE1</b>					
Between groups	.546	1	.546	.038	.846
Within groups	2613.133	180	14.517		
Total	2613.680	181			
<b>SE2</b>					
Between groups	3.219	1	3.219	.351	.554
Within groups	1643.277	179	9.180		
Total	1646.496	180			
<b>OE2</b>					
Between groups	16.725	1	16.725	1.599	.208
Within groups	1872.091	179	10.459		
Total	1888.816	180			
<b>SE3</b>					
Between groups	1.155	1	1.155	.226	.635
Within groups	925.569	181	5.114		
Total	926.724	182			
<b>OE3</b>					
Between groups	3.063	1	3.063	.382	.537
Within groups	1417.864	177	8.011		
Total	1420.926	178			
<b>SE4</b>					
Between groups	2.711	1	2.711	.486	.487
Within groups	1009.773	181	5.579		
Total	1012.484	182			
<b>OE4</b>					
Between groups	2.006	1	2.006	.658	.418
Within groups	551.732	181	3.048		
Total	553.738	182			
<b>SE5</b>					
Between groups	8.763	1	8.763	1.129	.289
Within groups	1405.132	181	7.763		
Total	1413.895	182			
<b>OE5</b>					
Between groups	27.681	1	27.681	2.616	.108
Within groups	1893.751	179	10.580		
Total	1921.431	180			

**Table 8**

*Analysis of Variance Results for Self-Efficacy (SE) and Outcome Expectancy (OE) by Section—Version 9*

	Sum of squares	<i>df</i>	Mean square	<i>F</i>	Sig
SE1					
Between groups	228.479	5	45.696	5.657	.000
Within groups	1445.799	179	8.077		
Total	1674.278	184			
OE1					
Between groups	200.483	5	40.097	3.138	.010
Within groups	2299.667	180	12.776		
Total	2500.150	185			
SE2					
Between groups	243.399	5	48.680	5.222	.000
Within groups	1687.444	181	9.323		
Total	1930.843	186			
OE2					
Between groups	178.812	5	35.762	4.055	.002
Within groups	1605.301	182	8.820		
Total	1784.113	187			
SE3					
Between groups	137.773	5	27.555	6.385	.000
Within groups	772.425	179	4.315		
Total	910.198	184			
OE3					
Between groups	123.086	5	24.617	3.685	.003
Within groups	1222.558	183	6.681		
Total	1345.644	188			
SE4					
Between groups	156.963	5	31.393	6.275	.000
Within groups	905.512	181	5.003		
Total	1062.475	186			
OE4					
Between groups	43.189	5	8.638	3.617	.004
Within groups	436.976	183	2.388		
Total	480.165	188			
SE5					
Between groups	87.889	5	17.578	2.242	.052
Within groups	1418.802	181	7.839		
Total	1506.691	186			
OE5					
Between groups	51.324	5	10.265	1.082	.372
Within groups	1698.769	179	9.490		
Total	1750.093	184			

**Table 9***Analysis of Variance Results for Self-Efficacy (SE) and Outcome Expectancy (OE) by Section—Version 8*

	Sum of squares	df	Mean square	F	Sig
<b>SE1</b>					
Between groups	14.688	5	2.938	.441	.819
Within groups	1184.638	178	6.655		
Total	1199.326	183			
<b>OE1</b>					
Between groups	92.964	5	18.593	1.298	.267
Within groups	2520.716	176	14.322		
Total	2613.680	181			
<b>SE2</b>					
Between groups	17.068	5	3.414	.367	.871
Within groups	1629.428	175	9.311		
Total	1646.496	180			
<b>OE2</b>					
Between groups	74.826	5	14.965	1.444	.211
Within groups	1813.990	175	10.366		
Total	1888.816	180			
<b>SE3</b>					
Between groups	47.242	5	9.448	1.902	.096
Within groups	879.482	177	4.969		
Total	926.724	182			
<b>OE3</b>					
Between groups	31.034	5	6.207	.773	.571
Within groups	1389.892	173	8.034		
Total	1420.926	178			
<b>SE4</b>					
Between groups	20.850	5	4.170	.744	.591
Within groups	991.634	177	5.602		
Total	1012.484	182			
<b>OE4</b>					
Between groups	8.261	5	1.652	.536	.749
Within groups	545.477	177	3.082		
Total	553.738	182			
<b>SE5</b>					
Between groups	69.719	5	13.944	1.836	.108
Within groups	1344.176	177	7.594		
Total	1413.895	182			
<b>OE5</b>					
Between groups	27.393	5	5.479	.506	.771
Within groups	1894.039	175	10.823		
Total	1921.431	180			

qualities for use with preservice elementary education teachers to assess self-efficacy in regard to the teaching of science as inquiry. Instrument reliability results are summarized in [Tables 2–5](#). The Cronbach’s alpha internal consistency data revealed an acceptable level of reliability in the scores for first generation instruments (Nunnally, 1978; Sax, 1974).

### **Implications**

Based on the study results, the experiences of the investigators, and prior literature, three areas of implications are specifically addressed. These areas are implications for research, policy, and practice.

#### **Research Implications**

The construct validity of an instrument is never fully established or achieved (Nunnally, 1970); thus, it is important to continue examining the construct validity of the TSI. In the process, the reliability of the instrument, including test-retest reliability, should continue to be assessed.

There are several research implications targeted toward teacher efficacy and science education. Research is needed to further explore the effects of self-efficacy on teacher development and how self-efficacy may affect eventual classroom practice. The TSI is a valuable tool for science teacher educators working in practical and research settings to assess the self-efficacy beliefs of prospective elementary teachers with regard to the teaching of science as inquiry.

The TSI should be used in combination with other data collection techniques to more fully determine the self-efficacy beliefs of prospective teachers. These data collection techniques may include, but are not limited to, observations of teachers engaged in the teaching of science as inquiry, as well as interviews with prospective teachers, to more clearly understand their ideas and beliefs associated with the teaching of science as inquiry. Although quantitative and qualitative research methods have been regarded as being “fundamentally different modes of inquiry,” both can be pursued rigorously (Shavelson & Towne, 2002, p. 19). Shavelson and Towne noted that the current trend of research in education to make greater use of qualitative methods at the expense of quantitative methods has created a dialogue of criticism. The nation’s commitment to make “. . .scientific literacy for all a reality in the 21st century” (NRC, 1996, p. ix) requires continued efforts to improve the research capacity of science education (NRC; Shavelson & Towne). “What makes research scientific is not the motive for carrying it out, but the manner in which it is carried out” (Shavelson & Towne, p. 20). Hence, using a mixed-methods approach to investigate preservice teacher self-efficacy should assist in achieving this goal.

Observation of preservice teachers in the classroom could provide additional information in relation to the predictive validity of the instrument. By observing classroom teaching, one would be able to determine if a particular score on the instrument transferred into behavior and practice. Additionally, development of a form of the TSI for practicing elementary teachers, as was done with the STEBI-B,

should be pursued. Furthermore, using the information obtained from both versions of this instrument, as well as classroom observations and interviews, comparisons can be made between the scores of preservice and inservice elementary science teachers.

In addition, because the instrument utilized a forced-choice response from the prospective elementary teachers, interviews should also be conducted. These interviews may more fully explore the prospective elementary teachers thoughts associated with the items on the TSI, as well as their responses to these items. This interview process could indicate if the preservice teacher truly understood the meaning of the items and if the researcher thoroughly understood the prospective elementary teacher's responses to the items. The inclusion of interviews would allow for a deeper understanding of the preservice teachers' self-efficacy in regard to the teaching of science as inquiry.

The idea of preservice elementary teachers' self-reporting inflated self-efficacy perceptions with regard to the teaching of science as inquiry also needs to be investigated. Because this study relied on self-reported data, it would behoove researchers to further investigate the idea of social desirability bias ([SDB]; Nancarrow & Brace, 2000; Ray, 1990). SDB refers to the possibility of respondents reporting what they perceive is socially desirable, rather than what might be the actual case. Such researchers as Phillips and Clancy (1972, as cited in Nancarrow & Brace, 2000) believed that SDB occurs because of two factors: "the general strength of need for approval felt by an individual (personality trait) and the demands of a particular situation." Another important factor that may contribute to SDB is the desire "to present oneself in a favorable light to others and/or a self-esteem preservation function" (ac cited in Nancarrow & Brace). Although these are possible concerns, there are many ways in which researchers may reduce this problem. For example, a Likert-type scale, similar to the one used for the TSI, is one viable solution to the problem of SDB (Ray).

Another area warranting further consideration is the concept of types of teacher efficacy. Although research has indicated that positive teacher efficacy is an appropriate goal (Ashton, 1985; Ashton & Webb, 1982, 1986; Bandura, 1977, 1986; Tschannen-Moran & Woolfolk-Hoy, 2001), Wheatley (2000) identified eight different types of positive teacher efficacy that he considers problematic: "traditional methods, traditional goals, too-certain efficacy, overly-optimistic novices, hypothetical future efficacy, pretend teacher efficacy, competitive teacher efficacy, and independent teacher control" (pp. 18–21).

The foundation for Wheatley's (2000, 2002) research is the belief that teachers' doubts about their teaching efficacy often have important benefits for teacher learning and education reform. Wheatley believed that these "doubts are essential to widespread success of education reform, particularly for reforms that promote progressive meaning-centered education" (2002, p. 5). Although these assertions conflict with most of the previous research on teacher efficacy, it is important to carefully explore the meaning of these findings, as well as their relationship to education reform. Thus, to more fully understand and encourage the types of teacher efficacy that support teacher development, new approaches to teacher efficacy

research are needed. To identify how teacher efficacy, confidence, and doubt may work together, research should be conducted within the context of the daily realities of teaching (2002). Additionally, these new approaches for investigating teacher efficacy should include qualitative means of research.

### Policy and Practice

The TSI is a useful tool for the evaluation of science method courses with an emphasis on inquiry teaching. Although research and evaluation are two separate entities, through the use of the aforementioned research techniques and methods, program improvement, development, and assessment should be facilitated.

“Like other research, evaluation attempts to describe, to understand the relationships between variables, and to trace out the causal sequence” (Weiss, 1972, p. 8). Evaluation applies the methods of social research and is distinguished by its intent, the purpose for which it is done, rather than the method or subject matter. The purpose of evaluation research is to measure the effects of a program against the goals it set out to accomplish (Payne, 1994; Weiss, 1972). For example, test-retest data derived from the TSI instrument can be used in combination with other research techniques using an experimental model. This evaluation can help to identify if a particular course is achieving what it purports with regard to the teaching of science as inquiry. The data derived from this analysis is then used to make decisions about the current program and future programs.

When analyzing this data, however, it is important for researchers to realize that reliability values may decrease from pretest to posttest. This, in fact, was evident in the reliability results presented in Tables 2 and 4. The scores reported for “learner engages in scientifically oriented questions” in relation to self-efficacy decreased from an alpha of .6884 on the pretest to an alpha of .6579 on the posttest. It is the researchers’ conjecture that when students initially entered the science methods course, their conceptions associated with learner engages in scientifically oriented questions were different than what is actually involved in the teaching of science as inquiry. Teaching science as inquiry requires teachers to possess a “sophisticated set of judgments about science, students, learning, and teaching” (NRC, 1996, p. 37). The central strategy for teaching science as inquiry is to use authentic questions generated from students’ experiences. Teachers provide students with the opportunity to investigate these questions by giving students investigations or by guiding students toward designing investigations of their own. As a result,

*Teachers of science are constantly making decisions, such as when to change the direction of a discussion, how to engage a particular student, when to let a student pursue a particular interest, and how to use an opportunity to model scientific skills and attitudes. (p. 33)*

Consequently, this complex decision-making process requires teachers to struggle with the tension between guiding students toward a set of predetermined goals and allowing students to set and meet their own goals (NRC, 1996). Thus, when preservice teachers have the opportunity to experience a science methods



course that provides preservice teachers the opportunity to experience the teaching of science as inquiry, their previously held conceptions may change. These preservice teachers may come to realize that the teaching of science as inquiry is much more complex and more difficult than they had originally thought.

### References

- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Anderson, O. (1997). A neurocognitive perspective on current learning theory and science instructional strategies. *Science Education, 81*, 67–89.
- Ashton, P. (1985). Motivation and the teacher's sense of efficacy. In C. Ames & R. Ames (Eds.), *Research on motivation in education: Vol. 2. The classroom milieu* (pp. 141–174). Orlando FL: Academic Press.
- Ashton, P. T., & Webb, R. B. (1982). *Teachers' sense of efficacy: Toward an ecological model*. Paper presented at the annual meeting of the American Education Research Association, New York.
- Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teacher efficacy and student achievement*. New York: Longman.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191–215.
- Bandura, A. (1981). Self-referent thought: A developmental analysis of self-efficacy. In J. H. Flavell & L. Ross (Eds.), *Social cognitive development frontiers and possible futures* (pp. 200–239). Melbourne, Australia: Cambridge University Press.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*, 122–147.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology, 25*, 729–735.
- Bandura, A. (1995). *Self-efficacy in changing societies*. Melbourne, Australia: Cambridge University Press.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman.
- Damnjanovic, A. (1999). Attitudes in inquiry-based teaching: Differences between preservice and inservice teachers. *School Science and Mathematics, 99*, 71–76.
- Drayton, B., & Falk, J. (2001). Tell-tale signs of the inquiry-oriented classroom. *NASSP Bulletin, 85*(623), 24–34.
- Enochs, L., & Riggs, I. (1990). Further development of an elementary science teaching efficacy belief instrument: A preservice elementary scale. *School Science and Mathematics, 90*, 694–706.
- Enochs, L., Scharmann, L., & Riggs, I. (1995). The relationship of pupil control to preservice elementary science teacher self-efficacy and outcome expectancy. *Science Education, 79*(1), 63–75.

- Freedman, M. (1997). Relationship among laboratory instruction, attitude toward science, and achievement in science knowledge. *Journal of Research in Science Teaching*, 34, 343–357.
- Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, 76, 569–582.
- Graziano, A. M., & Raulin, M. L. (2000). *Research methods: A process of inquiry* (4th ed.). Needham Heights, MA: Allyn and Bacon.
- Isaac, S., & Michael, W. B. (1997). *Handbook of research and evaluation*. San Diego, CA: Educational and Industrial Testing Services.
- Kleine, K., Brown, B., Harte, B., Hilson, A., Malone, D., Moller, K., et al. (2002). Examining inquiry. *Principal Leadership*, 3(3), 36–39.
- Lee, C., & Krapfl, L. (2002). Teaching as you would have the teach: An effective elementary science teacher preparation program. *Journal of Science Teacher Education*, 13, 247–265.
- Martin, (2000). *Elementary science methods: A constructivist approach*. Belmont, CA: Wadsworth Thomas Learning.
- Nancarrow, C., & Brace, I. (2000). Saying the “right thing”: Coping with social desirability bias in marketing research. *Bristol Business School Teaching and Research Review*, 3. Retrieved December 11, 2003, from [http://www.uwe.ac.uk/bbs/trr/Issue3/Is3-2\\_2.htm](http://www.uwe.ac.uk/bbs/trr/Issue3/Is3-2_2.htm).
- National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.
- National Research Council. (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. Washington, DC: National Academy Press.
- Nunnally, J. C. (1970). *Introduction to psychological measurement*. New York: McGraw-Hill.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw Hill.
- Payne, D. A. (1994). *Designing educational project and program evaluations: A practical overview based on research and experience*. Boston: Kluwer Academic Publishers.
- Ray, J. J. (1990). Acquiescence and problems with forced-choice scales. *Journal of Social Psychology*, 130, 397–399.
- Riggs, I. (1988). *The development of an elementary teachers' science teaching efficacy belief instrument*. Unpublished doctoral dissertation, Kansas State University, Manhattan.
- Riggs, I. (1998). *The impact of training and induction activities upon mentors as indicated through measurement of mentor self-efficacy*. Paper presented at the annual international conference of the Association for the Education of Teachers in Science, Minneapolis, MN.
- Riggs, I., & Enochs, L. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education*, 74, 625–637.
- Sax, G. (1974). *Principles of educational measurement and evaluation*. Belmont, CA: Wadsworth.

- Schwab, J. (1962). *The teaching of science: The teaching of science as enquiry*. Cambridge, MA: Harvard University Press.
- Shavelson, R. J., & Towne, L. (2002). *Scientific research in education*. Washington, DC: National Academy Press.
- Tschannen-Moran, M., Woolfolk-Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783–805.
- Von Secker, C. (2002). Effects of inquiry-based teacher practices on science excellence and equity. *The Journal of Educational Research, 95*, 151–160.
- Von Secker, C., & Lissitz, R. (1999). Estimating the impact of instructional practices on student achievement in science. *Journal of Research in Science Teaching, 36*, 1110–1126.
- Weiss, C. H. (1972). *Evaluation research: Methods of assessing program effectiveness*. Englewood Cliffs, NJ: Prentice-Hall.
- Weiss, I. R., Banilower, E. R., McMahon, K. C., & Smith, P. S. (2001). *Report of the 2000 national survey of science and mathematics education*. Chapel Hill, NC: Horizon Research.
- Wells, G. (1995). Language and the inquiry-oriented curriculum. *Curriculum Inquiry, 25*, 233–269.
- Wheatley, K. F. (2000). Positive teacher efficacy as an obstacle to educational reform. *Journal of Research and Development in Education, 34*(1), 14–27.
- Wheatley, K. F. (2002). The potential benefits of teacher efficacy doubts for educational reform. *Teaching and Teacher Education, 18*, 5–22.
- Windschitl, M. (2002). Inquiry projects in science teacher education: What can investigative experience reveal about teacher thinking and eventual classroom practice? *Science Teacher Education, 87*, 112–143.