5-2016

Investigating the Factor Structure of the Teachers' Sense of Efficacy Scale with Pakistani In-service and Pre-service Teachers

Sajid Ali Yousuf Zai
University of Arkansas, Fayetteville

Investigating the Factor Structure of the Teachers' Sense of Efficacy Scale with Pakistani In-service and Pre-service Teachers

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Educational Statistics and Research Methods

by

Sajid Ali Yousuf Zai
University of Sindh,
Bachelor in Computer and Information Technology, 2001
University of Sindh,
Master in Education, 2008

May 2016
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Dr. Ronna Turner
Dissertation Director

_____
Dr. Wen-juo Lo
Committee Member

_____
Dr. Chris Goering
Committee Member

_____
Dr. Parveen Munshi
Ex-Officio Member

Abstract

Teachers with a high sense of efficacy are motivated to achieve and are generally optimistic about future learning. There is an extensive body of research that indicates a teacher's self-efficacy beliefs can be a performance indicator for school outcomes. Research on characteristics related to teachers in Pakistan has been increasing over the last decade, however there are a number of instruments being used with this population without any documented validation studies. The purpose of this study was to investigate the validity of the Teacher Sense of Efficacy Scale (TSES; Tschannen-Moran & Woolfolk Hoy, 2001) to determine the latent structure of the TSES in the context of Pakistan in-service and pre-service teachers. Participants included 549 in-service (31% male and 69% female) and 423 pre-service (27% male and 73% female) teachers from four provinces of Pakistan. Content validity was investigated using experts' judgement ratings. All items were rated as culturally appropriate for a Pakistani population. This study evaluated the construct validity of the TSES using structure equation modeling. Three-factor models were hypothesized for in-service teachers and one-factor models for pre-service teachers, as had been consistent with factor structure of the TSES for in-service and pre-service teachers from other cultures. Confirmatory factor analysis validated the three-factor model for in-service teachers, as had been observed with other cultures. However, it did not support the one-factor model for pre-service teachers. As a follow-up, exploratory factor analysis produced three-factors for pre-service teacher, concluding that a three-factor model is more appropriate for both pre-service and in-service teachers in Pakistan. A multitrait-multimethod procedure provided partial evidence of convergent validity, however the scales within the TSES appear more correlated with each other than corresponding measures of the

scale. Teacher subgroup comparisons revealed that female teachers tend to have a higher sense of efficacy in student engagement, instructional strategies, and classroom management than male teachers. Moreover, Pakistani teachers teaching at primary level grades report a higher sense of efficacy than teachers teaching at higher grade levels. The findings of this study provide significant benefits for Pakistani researchers who want to use a teacher efficacy instrument as a tool for their studies.

Acknowledgment

I am grateful for the guidance and mentorship that Dr. Ronna Turner, my dissertation chair, has provided since the first day of my Ph. D in the University of Arkansas. I would not be able to complete this milestone timely without her support. I have learned so much from her over the past four years. She is a true inspiration for me.

I am grateful that I have improved both teaching skills and statistical skills under the guidance of Dr. Wen-juo Lo. His detail explanations and suggestion that "learn from the foundation" strengthened my comprehension and application of statistics. Thanks to Dr. Chris Goering, I really appreciate your time and efforts to be a part of the dissertation committee. Your valuable feedback improved the quality of my work. I am glad to have Dr. Parveen Munshi in my dissertation committee. Her support and encouragement to pursue doctorate degree in the field of statistics and research methods brought me in this department. Her expertise in Pakistani education system added more worth in my research. I am thankful to all my professors who taught, encouraged, and belief in me during my Ph.D.

I would also like to thank to the United States Agency for International Development (USAID) and Training for Pakistan (TFP) project for providing me the Ph.D and thanks to the World Learning for their continue assistance and logistics support throughout this Ph.D. journey.

Thank you so much to Dr. Jennifer Lynn for being a great friend and driving force during my Ph.D. I am also incredibly grateful for my program batch-mates, scholarship-mates and other American, Pakistani, and international friends who stood by to support me along the way.

Dedication

I would like to dedicate this dissertation to my family. This research could not have been done without my family who supported, appreciated, and prayed for me all the time. To my father, Fazal-e-Rabbi Yousuf Zai and my mother, my wife, my brothers: Jawaid Iqbal and family, Zafar Iqbal and family, Amjad Ali and family, and my sisters, who have been my encouragers and cheerleaders. I am overwhelmed and humbled by the never ending love, support, and kindness of my family and friends.

Table of Contents

List of Tables

List of Figures

CHAPTER I

Introduction

Education is the process of helping the up-coming generation acquire knowledge to build skills that allow them to interact successfully with their environment. The acquisition of competency and proficiency in instructional methods, student engagement, and classroom management are skills that are necessary for every teacher in order to facilitate the learning process. To see the extent and quality of a teacher's educational development through his / her level of performance requires a comprehensive program of evaluation. Evaluation is a process of assessing the attainments of the individual to satisfy the basic considerations underlying education. Although teacher evaluation processes are typically conducted by external evaluators, an internal belief that one can successfully utilize the training completed to facilitate learning is considered an important factor in effective teaching (Armor et al, 1976; Ashton & Webb, 1986; Henson, 2001a; Hollon, Anderson & Roth, 1991; Tschannen-Moran, Woolfolk Hoy & Hoy, 1998). Efficacy beliefs are judgments about one's perceived ability to carry out particular courses of action. In the teaching context, efficacy beliefs are formed from judgments about the difficulty of the teaching task and judgments about teaching ability (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998).

Many researchers have developed definitions of teacher efficacy and produced scales to measure teachers' efficacy levels (Fives & Buehl, 2010; Tschannen-Moran & Woolfolk Hoy, 2001). Tschannen-Moran and Woolfolk Hoy (2001) produced a scale to measure teachers' sense of efficacy for teaching that is currently being used by many educational researchers. This well-known teaching efficacy scale is named as the "Teachers' Sense of Efficacy Scale" (TSES). The

TSES has been validated for its use with both pre-service and in-service teachers from a number of countries including: United States (Tschannen-Moran & Woolfolk Hoy, 2001; Fives & Buehl, 2010), Europe (Klassen et al., 2009), Singapore (Klassen et al., 2009; Nie, Lau, & Liau, 2012), China, Korea, and Japan (Ruan et al., 2015).

Some of the previous researchers have used the translated versions of the TSES to validate the instrument for their specific populations (Al-Khalaileh, & Abu-Tineh, 2011; Çapa, Çakıroğlu, & Sarıkaya, 2005; De Stercke, Temperman, De Lièvre, & Lacocque, 2014; Guerreiro-Casanova, & Azzi, 2013; Tsigilis, Grammatikopoulos, & Koustelios, 2007). The instrument is also being used by researchers in some countries (i.e., Pakistan, India, Malaysia, Saudi Arabia) where the instrument has not been evaluated for the subpopulation. Specifically, researchers in Pakistan have begun using the TSES for studies with pre-service and in-service teachers without proper validation and review of the instrument's psychometric properties within Pakistan's context. It is therefore, the aim of this study to evaluate the construct validity of the TSES instrument in Pakistan.

Theoretical Framework

Teacher education has always been a central issue in the Pakistani Education system. Teachers' participation in decision-making and teachers' empowerment is considered to be an important factor or entity. Teachers influence students' achievement directly, and the teachers are influenced by their school leadership. Research indicates that the educators who make a difference in students' learning are led by head-teachers who make a significant and measurable contribution to the effectiveness of teachers and in the learning of pupils in their charge (Hallinger & Heck, 1998).

Teachers with a high sense of efficacy are motivated to achieve and are generally optimistic about future learning. Generally, teacher self-efficacy indicates teachers' beliefs in their ability to influence their students' achievement (Bandura, 1997; Skaalvik & Skaalvik, 2010). Bandura (1997) named this belief as perceived self-efficacy. A teacher's self-efficacy can also have an influence on the teacher's instructional practices and classroom management (Henson, 2001a; Hollon, Anderson & Roth, 1991) as well as student achievement (Armor et al, 1976; Ashton & Webb, 1986; Henson, 2001a).

There are two recognized theories that are used to explain and measure the concept of self-efficacy: Rotter's (1966) social learning theory and Bandura's (1977) social cognitive theory. The Rotter's (1966) social learning theory is based on internal verse external control beliefs. The internal control teacher is defined as teachers who believe that they have the abilities to teach more challenging and unmotivated students where the external control teachers believe that the outside class environment has more impact on students' learning than their own teaching (Rotter, 1966). Bandura's social cognitive theory (1977) is defined as individuals "function as contributors to their own motivation, behavior, and development within a network of reciprocally interacting influences" (Bandura, 1999, p. 169). Bandura's (1977) theory is the foundation of teacher efficacy and the theoretical foundation on which teacher efficacy is constructed. In recent years, the primary focus on cognitive theory is self-efficacy. Tschannen-Moran, Woolfolk Hoy and Hoy (1998) provide a definition of teachers' sense of self-efficacy as, "capabilities to organize and execute courses of action required to successfully accomplish a specific teaching task in a particular context" (p.233).

There have been several attempts made to measure teachers' efficacy using Rotter's (1966) Social Learning Theory (Armor et al., 1976; Ashton et al., 1982; Guskey, 1981; Rose & Medway, 1981) and Bandura's Social Cognitive Theory (Bandura, 1997; Gibson & Dembo, 1984; Tschannen-Moran & Woolfolk Hoy, 2001).

Tschannen-Moran and Woolfolk Hoy (2001) designed the TSES to measure teachers' sense of efficacy with three sub-scales: Student Engagement (SE), Classroom Management (CM), and Instructional Practices (IP). Originally, there were 52 items on the TSES but after pilot testing in three separate studies, the TSES was reduced to short (12-item) and long forms (24-item) of the measure. These scales focused on teachers' sense of efficacy as the belief in their competency to make a difference in students' performance (Tschannen-Moran & Woolfolk Hoy, 1998). Today, the TSES is widely used by educational and psychological researchers in different countries and it has been translated in different languages: Arabic (Al-Khalaileh & Abu-Tineh, 2011), Greek (Tsigilis, Grammatikopoulos, & Koustelios, 2007), Turkish (Çapa, Çakıroğlu & Sarıkaya, 2005), Chinese (Tschannen-Moran, n.d.), Portuguese (Guerreiro-Casanova & Azzi, 2013), and French (De Stercke, Temperman, De Lièvre, & Lacocque, 2014).

Statement of the Problem

Several research studies have been conducted to study teacher efficacy (Cantrell, Young & Moore, 2003; Plourde, 2002). However, in Pakistan, the research on teacher efficacy is in a relatively early stage. The TSES is used by a number of Pakistani researchers to examine the relationship between teacher self-efficacy and teachers' characteristics, such as teacher experience, teacher education, gender, school grade level, and student achievement (Ahmad, Khan, & Rehman, 2015; Butt, Khan & Jehan, 2012; Khan, 2012; Haq & Akhtar, 2013; Shaukat

& Iqbal, 2012). However, there is no published studies for the validation of the TSES with Pakistani teachers (either pre-service or in-service teachers). Thus, the findings of such research that use teacher's efficacy may be limited in its interpretation without a priori assessment of the use of this instrument with pre-service and in-service teachers from Pakistan.

Purpose of the Study

The purpose of this study is to investigate the validity of the Teacher's Sense of Efficacy Scale (TSES) and to determine the latent structure of the TSES in the context of Pakistani pre-service and in-service teachers. The study began with the foundational elements of item validity, scale reliability, and scale validity structure. This study also included construct validation through multi-trait multi-method analysis (Campbell & Fiske, 1959) comparisons, and include, examining the relationships between teachers' characteristics and their sense of efficacy level for scales demonstrating valid structure for the population.

Research Questions

The research questions that guided the researcher are divided into two sections with the validity process occurring first and second with descriptive analyses provided on the condition that the validity analysis provides sufficiently acceptable results.

Validation Process

1. Are items on the TSES interpreted differently by teachers in Pakistan as they are for the original population that the instrument was developed for?
2. Do the items function sufficiently within a scale-level internal consistency framework?

3. What is the teacher self-efficacy's latent structure for Pakistani pre-service and in-service teachers? Is the latent structure different for these two groups, similarly to data structures found in in-service and pre-service teachers in other populations?
4. Do the three sub-scales of TSES: Student Engagement (SE), Instructional Strategies (IS), and Classroom Management (CM) correlate as hypothesized with their convergent validity counterparts of alternative measures with same constructs in a multi-trait multi-method analysis?

Descriptive Comparison Study

5. What are the mean differences of the TSES subscales for in-service and pre-service teachers with varying background characteristics (e.g., gender, teaching experience, age, grade level taught, type of educational training, job status, and region of Pakistan)?
6. Is there a significant difference in pre-service teachers' efficacy beliefs based on gender and educational qualification?
7. Is there a significant difference in in-service teachers' efficacy beliefs based on gender, educational qualification, teaching experience, teaching grade level, and job status?

Significance of the Study

The study examined the validity and reliability of a teacher efficacy scale (Tschannen-Moran & Woolfolk Hoy, 2001) in Pakistan's context. A validation study has not been conducted using the TSES with this teacher population before, thus this study will have significant benefits for Pakistani researchers who use teacher efficacy instruments in their research. The results of this study provide an evidence of how factor structure can be influenced by different regions or

with different samples. Results may be provided that can assist efficacy experts with revising an instrument more appropriate for this specific population if needed.

Organization of the Study

This dissertation consists of five chapters. The first chapter of introduction includes the background of the study, theoretical framework, statement of the problem, purpose of the study, research questions, and significance of the study.

The chapter two of literature review provides an overview of theories used to define teacher efficacy belief (Bandura, 1977; Rotter, 1966), sources and factors of teacher efficacy, and measures of teacher's sense of efficacy. Several teacher efficacy scales that are commonly used by researchers are also discussed, along with a brief overview of the relationship between teacher efficacy and student achievement. Last, the Teacher Sense of Efficacy Scale (Tschannen-Moran & Woolfolk Hoy, 2001) and validation studies completed with this scale are presented.

The chapter three of methodology presents the research design. In this chapter, targeted participants and sample size are described. The instrument used in the current study are presented. In addition, data collection procedures are explained. Finally, statistical analysis procedures are described that are address the research questions.

Chapter four is a presentation of the instrument validations and group comparison results. Chapter five includes a discussion of the results, conclusions, limitations, and direction for future research.

CHAPTER II

Literature Review

This chapter provides a literature review of teacher efficacy theories, the factors that influence teacher efficacy, efficacy sources of information, and instruments to measure teacher efficacy. First, two foundation theories of efficacy will be discussed. Then a few of the most common teacher efficacy scales will be introduced along with their validation studies. Secondly, the development of the Teacher Sense of Efficacy Scale (Tschannen-Moran & Woolfolk Hoy, 2001) and its validation studies in different countries will be discussed. Finally, literature will be reviewed that assesses the relationship between teacher efficacy and student achievement, along with teacher demographics that are related to teacher efficacy.

Introduction of Teacher Efficacy

Being a key agent in the development of children, teachers play a pivotal role in the education system of a society. The success of an education system largely depends on the competence of teachers and their ability to implement effective pedagogies to facilitate student learning. Teacher education programs for pre-service teachers are designed to equip them with necessary pedagogical skills, instructional strategies, content knowledge, and the ability to communicate knowledge effectively.

Teacher efficacy has become a vital component of teacher competency being measured within the last five decades. Efficacy beliefs are judgments about the ability to carry out particular courses of action. In the teaching context, efficacy beliefs are formed from judgments about the difficulty of the teaching task and judgments about teaching ability (Tschannen-Moran,

Woolfolk Hoy, & Hoy, 1998). Teachers' sense of efficacy has been hypothesized to have an influencing impact on student achievement (Moore & Esselman, 1992; Ross, 1992) and student motivation (Midgley, Feldaufer, & Eccles, 1989). Teachers with high sense of efficacy are motivated to achieve, tend to persist, try new learning processes or tools, and are generally optimistic about the future.

There are several teacher efficacy instruments developed by researchers to measure the belief of teachers about their ability to complete certain teaching tasks (Bandura, 1997; Gibson & Dembo, 1984; RAND [Armor et al, 1976]; Tschannen-Moran & Woolfolk Hoy, 2001; [Webb Scale] Ashton et al., 1982). One of these measures has been widely used by educational researchers in many countries and in various languages, and is starting to be used with teachers in Pakistan even though there are no publically available validation studies for the Pakistani population. The purpose of this study is to examine the latent structure of Tschannen-Moran and Woolfolk Hoy's (2001) Teacher Sense of Efficacy Scale (TSES) for pre-service and in-service teachers in the context of Pakistan.

Teacher Sense of Efficacy

Tschannen-Moran, Woolfolk Hoy, and Hoy (1998) defined teacher efficacy as a teacher's beliefs about his/her competency to successfully complete a particular task related to his/her teaching. Teacher efficacy beliefs are beliefs that teachers hold about their skills to improve student performance in their classrooms. These beliefs have been found to be related to a wide range of positive outcomes for students, schools, teachers, and pre-service teachers (Tschannen-Moran, Woolfolk Hoy & Hoy, 1998). Some scholars suggest a causal and reciprocal effect in which higher perceived efficacy results in higher student achievement, which further strengthens

efficacy beliefs (Goddard, Hoy & Woolfolk Hoy, 2004). Rizvi and Elliot (2005) conducted a study to assess teachers' belief about their professional development. The factor analysis yielded two important factors of teacher efficacy which primary school teachers believe are essential for professional development. These dimensions are: belief that any given task can be achieved and belief in owning responsibility for student achievement (Rizvi & Elliot, 2005).

Personal efficacy is a teacher's self-confidence about his/her teaching abilities while general teaching efficacy relates to a general belief about his/her abilities of teaching to more challenging students (Woolfolk Hoy, 2000). These two constructs: personal efficacy and general efficacy are considered independent of each other (Hoy & Woolfolk, 1993).

Measure of Teacher's Sense of Efficacy

There are several attempts made to measure the self-efficacy of teachers. The two most common theoretical foundations of teacher self-efficacy are: Rotter's (1966) Social Learning Theory and Bandura's (1977) Social Cognitive Theory. As such, teacher efficacy scales are typically grounded in one of these two theories. Following are the most common teacher efficacy instruments used by researchers.

Rotter's Social Learning Theory (Rotter, 1966)

According to Rotter's Social Learning Theory (SLT), locus of control is categorized into two constructs: internal locus of control and external locus of control. In the internal locus of control, teacher efficacy is hypothesized as the level to which a teacher believes that his or her success or failure is due to his or her own ability. In contrast, external locus of control is referred to as a teacher level of efficacy belief that his or her success or failure is due to external factors

such as environment and luck (Rotter, 1966). The first recorded attempt to measure a teacher efficacy belief was carried out by a RAND (**R**esearch **an**d **D**evelopment; Armor et al, 1976) commissioned study.

The two-item RAND measure. The two-item teacher efficacy measure was developed by RAND on the basis of Rotter's (1966) internal and external locus of control theory. One item addressed internal locus of control and the other addressed external control. These two items are:

Rand item #1. *"When it comes right down on it, a teacher really can't do much because most of a student's motivation and performance depends on his or her home environment."*

Rand item #2. *"If I really try hard, I get through to even the most difficult or unmotivated students."*

The teachers' level of efficacy was determined by the sum of these two 5-point Likert scale items (Woolfolk Hoy & Burke–Spero, 2005). The aggregated response of these two-items was called teacher efficacy (TE). Item 1 was labeled as General Teaching Efficacy (GTE) and item 2 was labeled as Personal Teaching Efficacy (PTE). The researchers were concerned about the reliability of a two-item construct, and thus it encouraged other researchers to develop more reliable scales for comprehensive efficacy measurement (Tschannen-Moran & Woolfolk Hoy, 2001). Teachers' sense of efficacy and its relation to student achievement in reading scores was measured in the Rand study (Armor et al., 1976), along with a follow-up study by Armor et al. (1976) that assessed the degree to which teachers are confident that they have the skills to influence student achievement.

Responsibility of student achievement (Guskey, 1981). After the development of the two-item efficacy scale by Armor et al. (1976), Guskey (1981) developed a more wide-ranging teacher efficacy instrument called the Responsibility of Student Achievement (RSA) scale. The RSA is an efficacy instrument based on Rotter's internal and external locus of control. The RSA was developed to evaluate teachers' efficacy beliefs about their contribution to students' achievement. The RSA is a 30-item scale, and teachers are asked to respond to each item by distributing 100 percentage points between two alternative causal factors: whether success or failure of a student is due to the teacher's own action or whether the success or failure of a student is due to outside factors. An example item from this scale is:

If a student does well in your class, would it be

a. because that student had the natural ability to do well or

b. because of the encouragement you offered? (Guskey, 1981)

Guskey (1981) designed these two options based on Weiner's (1979) attributional theory. The RSA instrument measures three factors: teacher responsibility for student success (R+), teacher responsibility for student failure (R-), and the composite score of these two factors which is called responsibility of student achievement (Vasquez, 2009). There was a significant positive correlation between RSA and composite scores of teacher efficacy using RAND's two items in both student success (R+) and student failure (R-) (Guskey, 1982, 1988). This scale did not get attention from other researchers and currently no study has been found in the literature that adopted or tested this instrument other than the two Guskey studies (Guskey, 1982, 1988; Tschannen-Moran & Woolfolk Hoy, 2001).

Teacher Locus of Control (Rose & Medway, 1981). The Teacher Locus of Control (TLC) measure was developed by Rose and Medway (1981) in the same year as the Guskey (1981) study. This instrument is also based on Rotter's internal and external locus of control theory. The TLC instrument is based on 28-items in which half of the items describe situations of student success and the other half describes student failure situations. The format of the TLC is similar to Guskey's (1981) RSA instrument in which teachers were asked to provide responsibility of student success or failure by choosing two alternative options using a forced choice format. A sample item is:

When a student does better in school than he usually does, is it more likely

a. because the student was trying harder, or

*b. because you tried hard to encourage the student to do better? (*Rose & Medway, 1981, p. 189)

The 28-item TLC scale contains: 14 successes and 14 failure related items. Rose and Medway (1981) argued that the TLC scale was designed in this way because research indicated that locus of control can vary based on the trend of task performance result. In the original study, Rose and Medway (1981) found that there is a significant relationship between the TLC scale and student achievement. The author found the TLC has better reliability than Rotter's (1966) scale. When testing the validity of the TLC scale, Rose and Medway (1981) also found the TLC scale was a better predictor of teachers' behavior than Rotter (1966)'s internal / external scale (Tschannen-Moran & Woolfolk Hoy, 2001). The TLC scale was proposed as a valid instrument to measure teachers' belief regarding their control in the classroom.

The Webb Scale (Ashton et al., 1982). Another group of researchers (Ashton, Olejnik, Crocker, & McAuliffe, 1982) attempted to improve the teacher efficacy measure by reducing the problem of social bias. They did this by adding a forced-choice format with each item where social interest was considered present. This scale contains seven items. Participants must choose if they agree most strongly with the first or the second statement. A sample item is:

A. A teacher should not be expected to reach every child; some students are not going to make academic progress.
B. Every child is reachable. It is a teacher's obligation to see to it that every child makes academic progress.

Circle one:

1. I agree most strongly with A
2. I agree most strongly with B (Ashton et al., 1982).

There has been no study published on the *Webb Scale* (Ashton et al., 1982) beyond the original study up through 2015.

Bandura's Social Cognitive Theory (Bandura, 1977)

While the behavioral theorists of learning were of a view that learning is the result of associations formed by reinforcement, conditioning, and incentives, Albert Bandura added a social element and proposed the Social Cognitive Theory (SCT) with the idea that learning can be influenced by environment. Social Cognitive Theory provides a foundation for moral judgment and physiological stimulation, but researchers' primary focus on self-efficacy is the beliefs about an individual's confidence of successfully completing specific tasks (Locke &

Latham, 2002). Self-Efficacy is a part of the larger Social Cognitive Theory which is sometimes referred to as Social Learning Theory (SLT; Levin, Culkin, & Perrotto, 2001). Self-efficacy is one's belief that he or she can achieve success by his or her own skills under certain conditions. Bandura (1995) explains self-efficacy as "beliefs in one's capabilities to organize and execute the courses of action required to manage prospective situations" (p.2). In simple words, perceived self-efficacy can be defined as an individual's belief about his or her ability to get a successful outcome in a specific situation.

Teachers feel more comfortable and capable within their environment due to factors such as teacher training programs, experience, and academic attainment. Bandura (2006) stated that people are self-organizing, self-regulating, proactive, and self-reflecting. People will form intentions, set goals, anticipate outcomes, monitor actions, and reflect on personal efficacy. In learning situations, teachers may be able to self-reflect and establish the opportunity to set realistic goals for themselves and their students, and thus increases their perceived self-efficacy in the success of their teaching.

Sources of Self-Efficacy

According to Bandura (1997), the information regarding self-efficacy beliefs comes from four sources: *mastery experiences*, *learning through vicarious experiences*, *physiological and emotional states*, and *social persuasion*. Many researchers assess each source of self-efficacy independently (Gavora, 2010; Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998).

According to Bandura (1997), the most effective source for self-efficacy comes through mastery experiences or "*performance attainments*" (p. 399). Performance attainment is referred

to as a situation when a person experiences success in a given situation and he/she will have higher expectations of success if a similar situation occurs in the future while failure in a situation will lead to lower expectations in a similar future situation. The difficulty of the task and individual efforts also influence the development of self-efficacy belief.

The second source of self-efficacy development is vicarious experience. Vicarious experience is defined as an individual's judgments about his/her abilities to successfully perform a certain task based on the performance of other's and their experience (Bandura, 1997). This is indirect experience in developing self-efficacy. Observing someone who performed a task successfully suggests to the observer that he/she is capable of completing the task successfully. Similarly, an individual's self-efficacy is lowered when he/she observes that someone has failed to complete a task. According to Usher and Panjares' (2009) validation study, vicarious experience is a difficult factor to measure using traditional quantitative measures.

The third source of self-efficacy, physiological and emotional states, also influence the development of self-efficacy. Physiological and emotional states such as anxiety and stress effect one's assessment of his/her abilities (Bandura, 1997). Stress is a "sign of vulnerability to poor performance" (Bandura, 1995, p. 4). Psychological and emotional stress has also been used to assess students' anxiety (Usher & Panjares, 2009).

The fourth source that influences the development of self-efficacy is social persuasion. Social persuasion can be defined as verbal feedback or encouragement from others about one's performance. According to Bandura (1995, 1997), when an individual gets motivation or influence from others, his/her sense of efficacy is increased and he/she is more likely to put in effort to complete a certain task.

Figure 1. Sources of self-efficacy (Bandura, 1997)

Usher and Pajares's (2009) study developed and validated an instrument that measures Bandura's (1997) four hypothesized sources of self-efficacy. They found that vicarious experience is hard to measure, and the relationship between these four sources and self-efficacy should not be generalized (Usher & Pajares, 2009).

There have been several attempts made to construct a teacher efficacy scale since Bandura's (1977) social cognitive theory was introduced. Following are some of the most useful teacher efficacy measures.

The Ashton Vignettes (Ashton et al., 1984). Ashton, Buhr, and Crocker (1984) developed the first teacher efficacy scale based on Bandura's (1977) theory. This scale was based on 50 items with six dimensions: discipline, academic instruction, motivation and work

with parents, planning, and evaluation. In their scale, teachers were asked to rate his or her belief on a 7-point Likert scale from 1 (extremely ineffective) to 7 (extremely effective) in the given situation. A sample item is:

> *"This year your principal has assigned you to teach a class of low ability students in your subject matter area. The teacher who taught this class last year tells you that these are the slowest students that she's taught in her twenty year teaching career. How effective would you be in increasing the academic achievement of the students in this class?"* (Ashton et al., 1984)

Teacher Efficacy Scale (TES) (Gibson & Dembo, 1984). Gibson and Dembo's (1984) development of a Teacher Efficacy Scale (TES) included 208 elementary teachers completing a survey consisting of 30 items. Each item was on a 6-point Likert scale from "strongly disagree" to "strongly agree" to assess teachers' sense of efficacy. The factor loadings indicated that 16 items significantly loaded into one of two factors. The reliability for these 16 items on Gibson and Dembo's (1984) Teacher Efficacy Scale (TES) was .79, while the reliability coefficient for the first factor (Personal Teaching Efficacy) was .78 and for the second factor (Teaching Efficacy) was .75. A factor analysis was employed for interpreting the grouping of items and yielded two factors. The first factor was classified as teachers' sense of personal teaching efficacy, and the second factor was labeled teachers' sense of teaching efficacy. Since verbal ability and teacher flexibility is associated with teacher behavior and student achievement, Gibson and Dembo (1984) further investigated convergent and discriminant validity of teacher efficacy scale when compared with two other measures: verbal ability (Coleman et al., 1966) and flexibility (Ekstrom, 1975). Verbal ability was measured by the Verbal Facility Test (Coleman et al., 1966) and Controlled Associations Test (French, Ekstrom, & Price, 1963) while flexibility was measured using Finding Useful Parts and the Planning Test (French et al., 1963). The convergent validity coefficients for the three traits of teacher efficacy, verbal ability, and

flexibility were statistically significant although modest in magnitude, .42, .30, and .39, $p < .05$, respectively (Gibson & Dembo 1984).

This TES (Gibson & Dembo, 1984) scale has been widely used by many international researchers for their subpopulation despite the fact this instrument has some serious psychometrics problems. For example, there are inconsistencies in the factors of the instrument, with several items loading on multiple factors, low reliability for the subscales, and the lack of clarity about the operation definition of the two (i.e., PTE, GTE) constructs (Gavora, 2010; Soodak & Podell, 1996; Tschannen-Moran et al., 1998; Woolfolk Hoy & Burke-Spero, 2005).

Coladarci (1992) investigated the relationship between teacher self-efficacy and teacher commitment to teaching. Gibson and Dembo's (1984) TES teacher efficacy instrument was used for 170 elementary school teachers in Maine. Result showed significant, but small relationships of commitment to teaching with personal efficacy ($r = .25$), as well as with general efficacy ($r = .31$).

General Self-Efficacy (GSE) scale**.** This 10-item scale using a 4-point Likert scale format (1= not at all true, 2= hardly true, 3= moderately true, 4= exactly true) was originally developed by Matthias Jerusalem and Ralf Schwarzer in 1979 in a German version and was later translated into 31 other languages by various authors (Schwarzer & Jerusalem, 1995). This scale was designed to assess a general sense of perceived self-efficacy for the general adult population (Schwarzer & Jerusalem, 1995). The reliability coefficient for this unidimensional scale administered to adults in 23 countries ranged from .76 to .90, with a mode of .80. Sample items for the General Self-Efficacy (GSE) scale are:

- *"I can always manage to solve difficult problems if I try hard enough."* and
- *"I am confident that I could deal efficiently with unexpected events"* (Schwarzer & Jerusalem, 1995).

This efficacy scale is included in the review because of its use with a wide range of teacher populations (Ebstrup et al., 2011; Luszczynska et al., 2005). Ebstrup, Eplov, Pisinger, and Jørgensen (2011) attempted to evaluate the relationship between general self-efficacy and five personality-stress factors with 3,471 adults aged 18-69 in Denmark. The results suggested that all five personality traits: extraversion, agreeableness, conscientiousness, emotional stability, and openness were mediated by efficacy beliefs. Their results suggested that the GSE is an important mediator to consider in the relationship between personality and perceived stress of teachers (Ebstrup et al., 2011). A similar study was conducted in China and results were consistent with previous studies (Wang et al., 2014). Research indicates that General Self-Efficacy is significantly related to personality factors (i.e., optimism, self-regulation, self-esteem, and orientation towards the future) in five countries: USA, Costa Rica, Germany, Turkey, and Poland (Luszczynska, et al., 2005). This scale is available in 31 languages and the original author provided the data of almost 18,000 participants' responses from 24 countries on his official website (Schwarzer & Jerusalem, 2014).

The Norwegian Teacher Self-Efficacy Scale. The Norwegian Teacher Self-Efficacy Scale (NTSES) was developed by Skaalvik and Skaalvik (2010). The NTSES is composed of 24-items using a 7-point Likert-scale ranging from 1 (not certain at all) to 7 (absolutely certain). The NTSES consists of six factors with four items in each factor. This NTSES was developed under Bandura (1997)'s recommendations for efficacy construction. Skaalvik and Skaalvik (2010)

included 2249 teachers from 113 elementary and middle schools in Norway. The Cronbach's alphas for the six constructs along with sample items are:

1. *Instruction* (α = .83): "How certain are you that you can explain central themes in your subjects so that even the low-achieving students understand?"
2. *Adapt instruction to individual needs* (α = .90): "How certain are you that you can provide realistic challenges for all students even in mixed ability classes?"
3. *Cooperate with colleagues and parents* (α = .83): "How certain are you that you can cooperate well with most parents?"
4. *Cope with change* (α = .91): "How certain are you that you can successfully use any instructional method that the school decides to use?"
5. *Motivate students* (α = .77): "How certain are you that you can get all students in class to work hard with their schoolwork?"
6. *Maintain discipline* (α = .81): "How certain are you that you can maintain discipline in any school class or group of students?"

The NTSES has been used in a number of studies (Al-Alwan & Mahasneh, 2014; Avanzi et. el, 2013) and has been significantly related to job satisfaction, job burn-out (Avanzi, et. al, 2013) and students' attitude toward school (Al-Alwan & Mahasneh, 2014).

Bandura's Teacher Self-Efficacy Scale (Bandura, 2006). Bandura's updated Teacher Self-Efficacy Scale (Bandura, 2006) is 30 items on a 9-point Likert scale fixed as 1 = nothing, 3 = very little, 5 = some influence, 7 = quite a bite, and 9 = a great deal. This scale is originally updated from Bandura's (1997) work. The reliability of Bandura's teacher self-efficacy scale composite 30-item scale is .94 while the reliabilities for the seven subscales are: *instructional*

self-efficacy (α = .85), *disciplinary efficacy* (α = .90), *influence on decision making* (α = .86), *influence on school resources, enlisting parental involvement* (α = .81), *enlisting community involvement* (α = .87), and *creating a positive school climate* (α = .87) (Bandura, unpublished, undated; Quinn, 2007). Sample items from Bandura's (2006) efficacy scale are:

- *How much can you do to make students enjoy coming to class?*
- *How much can you do to get children to follow classroom rules?*

Billheimer (2006) investigated the efficacy differences between early childhood and elementary pre-school teachers using Bandura's efficacy scale. He found significant differences between early childhood and elementary education pre-school teachers in influencing decision making and creating a positive school climate subscales. Pre-school teachers of early childhood education tend to have higher efficacy than elementary education teachers on the influencing decision making subscale, t (86) = 3.36, $p < .05$; and on creating a positive school climate, t (86) = 3.01, $p < .01$; however there was no significance difference between groups found in total composite efficacy scores, t (86) = 1.44 (Billheimer, 2006; no probability value provided).

Quinn (2007) evaluated the relationship between Bandura's Teaching Self-Efficacy Scale (Bandura, 2006) and pre-school teachers. There were 88 pre-school and kindergarten teachers who participated in this study. Several demographic variables were tested to assess the relationship of efficacy with teacher demographic characteristics. Only educational degree was a significant predictor, F (1, 86) = 12.6, $p < .001$ with a 1-unit change in educational degree accounting for a 9.04 units change on the Teaching Self-Efficacy Scale (Bandura, 2006) scores holding other variables constant (Quinn, 2007). Bandura's (1997) teacher efficacy scale provided

the groundwork for the Teacher Sense of Efficacy Scale (Tschannen-Moran & Woolfolk Hoy, 2001).

Teacher Sense of Efficacy Scale (Tschannen-Moran & Woolfolk Hoy, 2001). After reviewing Bandura's (1997) efficacy scale and adding more items, Tschannen-Moran and Woolfolk Hoy (2001) developed one of the most recent and commonly used teacher efficacy scales currently in use. Tschannen-Moran and Woolfolk Hoy (2001) added items to measure the essential elements of effective teaching which they believed were missing in previous efficacy scales. The Teacher Sense of Efficacy Scale (TSES) which is also referred to as the *Ohio State Teacher Efficacy Scale* (OSTES), is a 24-item scale with three factors: student engagement, classroom management, and instructional strategies on a 9-point Likert scale similar to Bandura's TSS scale. Tschannen-Moran and Woolfolk Hoy (2001) produced two forms for the TSES: a long-form (24-item) with eight items per factor and a short-form (12-item) with four items per factor. The reliabilities for the composites TSES long-form and short-form are: .94 and .90 respectively. The TSES is widely used by researchers in different educational settings. This scale is also validated for many subpopulations; however, there are no validation studies published for Pakistani in-service and pre-service teachers. Therefore, the purpose of this study is to validate the TSES for in-service and pre-service teachers from Pakistan. A detailed review about the development of the TSES, its validation in different countries, and the relationship of the teacher efficacy with teacher characteristics are discussed in the following sections.

The Development of Teacher Sense of Efficacy Scale

The Teacher Self Efficacy Scale (TSES) which is also named the *Ohio State Teacher Efficacy Scale (OSTES)* was developed by a team of 10 members including two faculty

researchers and eight graduate students. Each member independently selected items from Bandura's scale that related to important teaching tasks. In addition, each member further created 8-10 new items related to teaching tasks which were missing in Bandura's scale. There were more than 100 items at this stage. Then the team selected 52 items after a thorough discussion, revision, and elimination of duplicated or overlapping items.

These 52 items represented a range of teaching tasks (Tschannen-Moran & Woolfolk Hoy, 2001). In this pool of 52 items, 23 items were adopted from Bandura's 30-item scale while the rest of the 29 items were new and generated by the group. These items represented important teaching related tasks which are not provided on the Bandura scale, such as classroom assessment, coping with unmotivated students, using flexible instructional strategies according to individual students' need, dealing with students' classroom behavior, and motivating students' classroom engagement (Tschannen-Moran & Woolfolk Hoy, 2001). These 52 items were placed on a 9-point Likert scale ranging from 1 = nothing, 3 = very little, 5 = some influence, 7 = quite a bit, and 9 = a great deal, similar to the Bandura's TSS scale. Three separate studies were conducted within the larger validity study to evaluate the reliability and validity of the TSES and further refine the scales. Tschannen-Moran and Woolfolk Hoy (2001) used the empirical approach of construct development for their selection of items. There are two main approaches of construct development: a rational approach and an empirical approach (Dawis, 1987). The rational approach (aka: theoretical approach) is an approach where operational definitions of scales are developed based on a theoretical framework, and items are developed to reflect the construct they are attempting to measure. Conversely, the empirical approach depends on data and statistical procedures to identify scale constructs based on the grouping of items (Dawis, 1987; Janda, 1998).

Tschannen-Moran and Woolfolk Hoy (2001) study 1. In the first study, the 52-item pool selected by the Tschannen-Moran and Woolfolk Hoy research team was reduced to 32 items based on their statistical analyses. There were 224 teachers from Ohio State University who participated in the first study including 146 pre-service teachers and 78 in-service teachers. Respondents were asked to rate each item on a 4-point scale (1 = not at all, 2 = somewhat, 3 = important, 4 = critical) instead of the 9-point scale described above (Tschannen-Moran & Woolfolk Hoy, 2001). The 52-item pool yielded ten factors with eigenvalues greater than one (Kaiser's rule, 1960), accounting for 57.2% of the variance using principal-axis factoring with a varimax rotation. Tschannen-Moran and Woolfolk Hoy (2001) set the criteria of .60 as a significant factor loading. This resulted in 31 items with factor loadings ranging from .62 to .78, with one item with a .595 loading also included because it reflected the important area of motivation. Thus, they reduced the 52-item pool into 32 items for further study.

Tschannen-Moran and Woolfolk Hoy (2001) study 2. A second study was conducted with a different sample using the refined 32-item pool that resulted in a final set of 18 items. The sample in the second study consisted of 217 participants including 70 pre-service teachers and 147 in-service teachers from three U.S. universities. The principal-axis factoring using a varimax rotation for the 32 items produced eight factors with eigenvalues greater than one, accounting for 63% of the variance. However, the scree plot suggested two or three factors (Tschannen-Moran & Woolfolk Hoy, 2001). In the three-factor solution, classroom management appeared as a separate factor along with two other factors of instructional strategies and student engagement. Since classroom management is an important aspect in effective teaching (Brophy & Good, 1986), Tschannen-Moran and Woolfolk Hoy (2001) decided to retain a three-factor solution. The 32-item pool was further reduced to 18 items after removing items with weak loadings, items

that loaded significantly on a factor other than the hypothesized factor, and overlapping items in the three factor model.

Tschannen-Moran and Woolfolk Hoy (2001) reanalyzed the 18-items using principal axis factoring with an orthogonal rotation. The results indicated a three-factor solution accounting for 51% of the total variance. This three factor result was consistent with previous results, as would be expected given the same data were used for the second analysis. These three factors were labeled: *Efficacy for Student Engagement* (8 items), *Efficacy for Instructional Strategies* (7 items), and *Efficacy for Classroom Management* (3 items).

Tschannen-Moran and Woolfolk Hoy's (2001) scale resulted in a three factor model based on their empirical approach, however they did not provide operational definitions for these factors. The reliability of the three subscales: *Efficacy for Student Engagement*, *Efficacy for Instructional Strategies*, and *Efficacy for Classroom Management* were .82, .81, and .72, respectively. The classroom management subscale did not produce sufficient reliability due to a small number of items (3 items). A second-order factor analysis was used to re-analyze three teacher efficacy subscales using the combined samples of study 1 and study 2. Principal axis factoring yielded one substantial factor with factor loadings ranging from .74 to 0.84. The authors concluded that this result suggested that the 18 items of teacher efficacy could be used as a unidimensional scale, as well as three subscales. To provide further evidence of the essential unidimensionality of a combined teacher efficacy scale, Tschannen-Moran and Woolfolk Hoy (2001) reanalyzed the data with one factor specification for the 18-item teacher efficacy scale using principal-axis factoring. All 18 items loaded on one factor with factor loadings ranging

from .48 to .70. The internal consistency reliability of the 18 items using the combined samples of both studies was .95.

The construct validity of the 18-item efficacy scale was assessed by correlating it and its three subscales with existing efficacy scales. The total scores of the 18-item efficacy scale were positively, but weakly correlated with both the general teaching efficacy RAND item #1 ($r = .35$, $p < .01$) and Gibson and Dembo's General Teaching Efficacy (GTE) subscale ($r = .30$, $p < .01$). The 18-item Tschannen-Moran and Woolfolk Hoy scale was also significantly and weekly correlated with the personal teaching efficacy RAND item #2 ($r = .28$, $p < .01$) and Gibson and moderately correlated with Dembo's Personal Teaching Efficacy (PTE) subscale ($r = .48$, $p < .01$). The author did not provide the TSES subscale correlations with the comparisons scales.

Tschannen-Moran and Woolfolk Hoy (2001) study 3. In the third study, Tschannen-Moran and Woolfolk Hoy (2001) further refined the teacher self-efficacy scale. They found that the classroom management subscale produced weak stability (Robert & Henson, 2001a) due to a few number of items. Therefore, the team decided to construct more items for the classroom management factor. They adopted items from Emmer's (1990) classroom management scale. In addition, they created more items that reflected teachers' tasks related to classroom management which were not available in Emmer's (1990)'s classroom management scale. The instrument was increased from 18 items scale to 36 items after including new classroom management items that were field-tested in a class and evaluated by the researchers. The sample for the third study consisted of 410 participants including 103 pre-service and 255 in-service teachers from three U.S. universities.

Principal axis factoring with a varimax rotation for the 36-item scale produced four factors using Kaiser's (1960) rule, accounting for 58% of the variance while the scree plot suggested three factors (Tschannen-Moran & Woolfolk Hoy, 2001). These three factors were labeled: Efficacy for Student Engagement (12 items), Efficacy for Instructional Strategies (15 items), and Efficacy for Classroom Management (9 items). Tschannen-Moran and Woolfolk Hoy (2001) reduced the 36 items into a 24-item scale by selecting the eight items with the highest loading values on each factor. The final version of the 24-item scale was reanalyzed using a principal axis factoring and unsurprisingly produced the same three factors with loadings ranging from .50 to .78.

The internal consistency reliability of the 24-item composite teacher efficacy scale was .94 and the reliabilities of the three subscales: student engagement (8-items), instructional strategies (8-items), and classroom management (8-items) were .87, .91, and .90, respectively. The results indicated large positive inter-correlations among the three subscale factors ranging from .58 to .70.

Further analyses revealed that the factor structure for pre-service teachers did not support the three-factor structure as was observed for in-service teachers. Therefore, Tschannen-Moran and Woolfolk Hoy (2001) recommended use of a single factor scale for pre-service teachers. They also tested both the long version (24-items) and short version (12-items) of the scale for only pre-service teachers' responses and they found 57% of the variance accounted for in a one-factor solution for the long version and 61% of the variance accounted for in a one-factor solution for short version with factor loadings ranging from .60 to .85. The authors concluded that total scores of the teacher efficacy scale are more appropriate and subscale scores are less

meaningful for pre-service teachers who have not yet experienced teaching-related tasks. The complete process of the development of the Teacher's Sense of Efficacy Scale (TSES) is illustrated in figure 2.

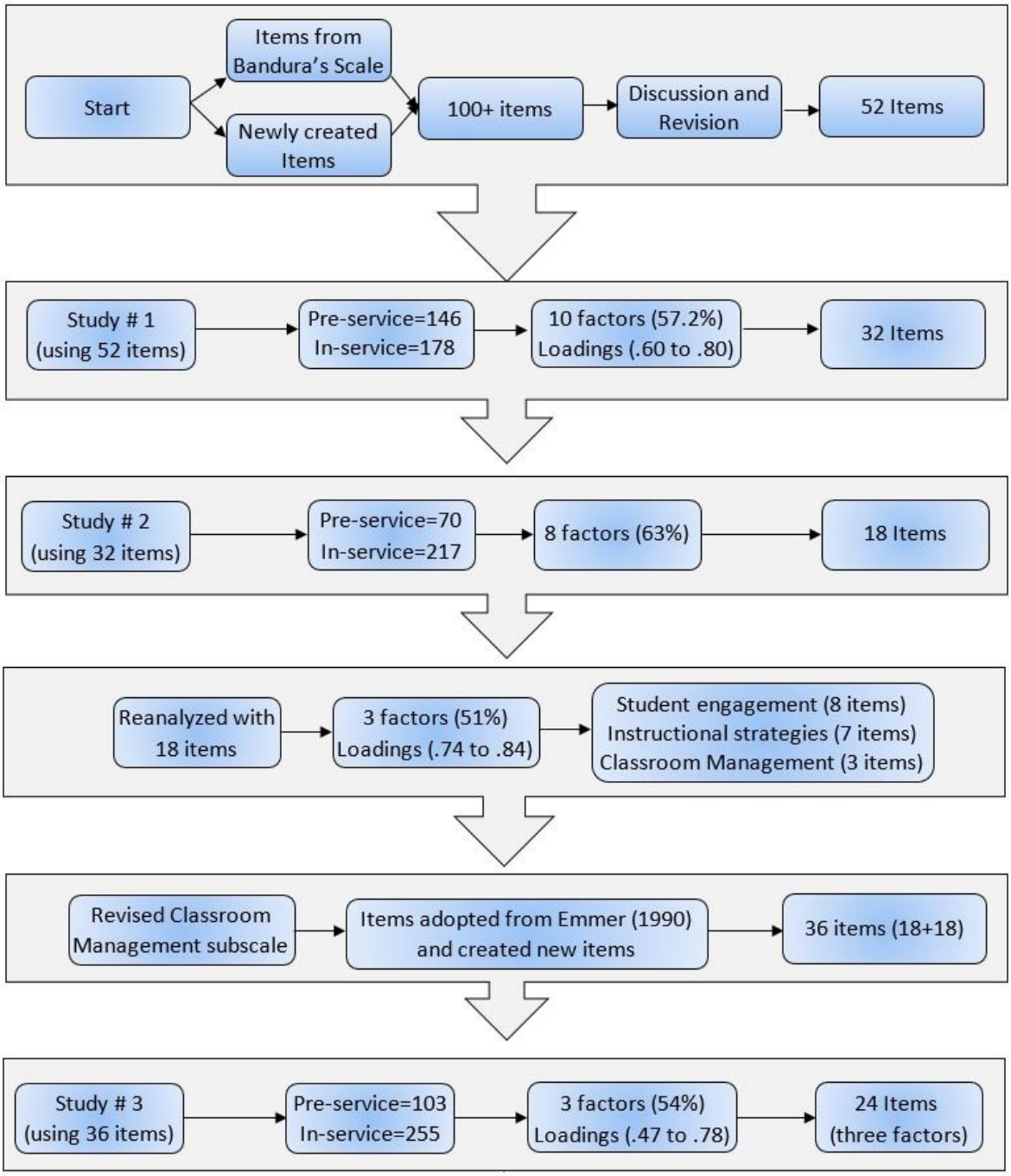

Figure 2. Development of the 24-item Teacher Self-Efficacy Scale (Tschannen-Moran & Woolfolk Hoy, 2001)

The construct validity of the TSES was also assessed by examining its relationship with existing efficacy scales. Table 1 (Tschannen-Moran & Woolfolk Hoy, 2001, p. 802) shows the correlation matrix of the TSES with existing efficacy measures.

Table 1

Correlation among the TSES Scale with Existing Efficacy Scales

	1	2	3	4	5	6	7	8
	TSES	Instruction	Management	Engagement	Rand 1	Rand 2	GTE	PTE
1		0.89**	0.84**	0.87**	0.18**	0.53**	0.16**	0.64**
2			0.60**	0.70**	0.07	0.45**	0.06	0.62**
3				0.58**	0.29**	0.46**	0.30**	0.45**
4					0.11*	0.47**	0.06	0.58**
5						0.23**	0.65**	0.12*
6							0.13*	0.65**
7								0.07

Source: Tschannen-Moran and Woolfolk Hoy (2001), p.802

* $p < .05$ (2-tailed); ** $p < .01$ (2-tailed); TSES = Teacher Sense of Efficacy (Tschannen-Moran & Woolfolk Hoy, 2001); GTE= General Teacher Efficacy (Gibson & Dembo scale); PTE= Personal Teacher Efficacy (Gibson & Dembo scale).

There are moderate to strong positive correlations among the three subscales of the TSES (.58 to .70). The total teacher sense of efficacy scale is significantly correlated with the PTE (Personal Teaching Efficacy) scale by Gibson and Dembo's (1984) and Rand (item #2). The TSES subscales of instructional strategies, classroom management, and student engagement are moderately and approximately equally correlated with the Rand #2 external locus of control item (correlation from .45 to .47) and slightly more correlated with the PTE scale with correlations ranging from .45 to .62. The two TSES scales of instructional strategies and student engagement have close to a zero relationship with the Rand # 1 internal locus of control item and the Gibson and Dembo (1984) GTE scale. The TSES classroom management scale has slightly larger, but

small relationships with the Rand #1 and the GTE scale indicating very little overlap in the constructs being measured.

Literature suggests that the TSES is currently used by a wide range of researchers for studying diverse populations. Researchers are also conducting validation studies of this instrument on specific populations. Currently, the TSES is considered one of the most reliable and evaluated instruments to measure the efficacy beliefs of pre-service and in-service teachers. The TSES has been evaluated several times with different subpopulations and it yields three factors: student engagement, instructional strategies, and classroom management (Tschannen-Moran & Woolfolk Hoy, 2001), most specifically for in-service teachers. Tschannen-Moran and Woolfolk Hoy believe these three factors represent requirements for an effective teacher. Research related to each of these scales is highlighted separately below.

Teacher efficacy for student engagement. Research indicates that teachers' practices influence student engagement (Skinner & Belmont, 1993). A study was conducted by Skinner, Wellborn, and Connell (1990) to evaluate the relationship between teachers' behavior and student engagement in the classroom. They found that teachers' behavior in the classroom is a significant predictor for students' engagement ($r^2 = .53$). Furthermore, student engagement is one of the factors related to higher scores on standardized academic tests (Skinner & Belmont, 1993). Efficacy for student engagement on the long form of the TSES is measured by eight items. Some of the sample items are:

- *How much can you do to motivate students who show low interest in school work?*
- *How much can you do to get students to believe they can do well in school work?*
- *How much can you do to foster student creativity?*

- *How much can you do to improve the understanding of a student who is failing?* (Tschannen-Moran & Woolfolk Hoy, 2001, p. 800)

Teacher efficacy for instructional strategies. The efficacy for instructional strategies is referred to as a teacher's confidence that he or she can design and implement classroom's activities, instructional styles, and assessment according to the needs of individual students (Wolters & Daugherty, 2007). Empirical evidence indicates that high levels of teacher efficacy in the area of instructional strategies is associated with students' motivation to learn (Mojavezi & Tamiz, 2012) and students' achievement (Wolters & Daugherty, 2007). Guskey (1988) conducted a study to assess the relationship between teacher efficacy belief and teachers' willingness to learn and the utilization of new instructional strategies. Teacher efficacy was measured using the RSA scale (Guskey, 1981) and two RAND items. The results of the Guskey (1988) study concluded that teachers with high levels of teaching efficacy are more willing to implement new instructional strategies in their teaching. Similar findings were revealed by Minke, Bear, Deemer and Griffen (1996). Saklofske, Michayluk and Randhawa (1988) also found there is significant correlation between personal teaching efficacy belief and classroom management behaviors ($r = .23$), however this relationship was relatively small. Wolters and Daugherty (2007) found a significant relationship between teachers' self-efficacy for instructional strategies and mastery goal structure ($r = .38$, $p < .001$). When school grades and teaching experience was added in the model, Wolters and Daugherty's (2007) study found that self-efficacy for instructional strategies using the TSES was a significant predictor ($\beta = .31$, $p < .001$) of mastery structure. Researchers have concluded that teachers' sense of efficacy influences instructional practices in the classrooms (Caprara, Barbaranelli, Steca, & Malone, 2006).

Teacher efficacy for instructional strategies on the long form of the TSES is measured by eight items. Following are some items that measure the instructional strategies construct:

- *How well can you respond to difficult questions from your students?*
- *How much can you do to adjust your lessons to the proper level for individual students?*
- *To what extent can you provide an alternative explanation or example when students are confused?*
- *How well can you provide appropriate challenges for very capable students*?

(Tschannen-Moran & Woolfolk Hoy, 2001, p. 800)

These questions relate to instructional strategies, coping with challenging students, and challenging more capable students. Teachers with high efficacy are believed to use more instructional methods to teach each student than teachers with a low level of efficacy (Ashton & Webb, 1986).

Teacher efficacy for classroom management. Classroom management is an important element in the learning process. Classroom management is also an essential component in effective teaching (Brophy & Good, 1986; Tschannen-Moran & Woolfolk Hoy, 2001). The management of the behavior of students and its implications for learning are an emerging concern for teachers, parents, and policy makers (Barker, Yeung, Dobia, & Mooney, 2009). A well-equipped teacher knows how to manage his/her class effectively. Students' disruptive behavior does not only impact their own learning outcomes, but it also negatively affects a teacher's wellbeing and self-efficacy (Lewis & Sugai, 1999). Teachers with high efficacy have been found to use alternative classroom management strategies for effective teaching (Ashton & Webb, 1986).

The teacher efficacy for classroom management subscale in the TSES is the revised updated version of Emmer (1990)'s classroom management scale. Tschannen-Moran and Woolfolk Hoy (2011) also included some items to measure additional aspects of teaching that have not been available in other teacher efficacy scales. The results using factor analysis procedures concluded that classroom management is a separate factor from the other two subtypes of efficacy (instructional strategies and student engagement). Teacher efficacy for classroom management on the long form of the TSES is measured by eight items. Some of the sample items are below:

- *How much can you do to control disruptive behavior in the classroom?*
- *How much can you do to get children to follow classroom rules?*
- *How well can you establish a classroom management system with each group of students?*
- *How well can you keep a few problem students form ruining an entire lesson?*

(Tschannen-Moran & Woolfolk Hoy, 2001, p. 800)

According to Allinder (1994), teachers with high sense of efficacy show greater levels of classroom management. Teachers with high level of efficacy can utilize classroom management strategies that can effectively address students with disruptive behavior compared to teachers with lower levels of efficacy (Morris-Rothschild & Brassard, 2006). Morris-Rothschild and Brassard (2006) assessed the relationship between teacher efficacy for classroom management and teacher conflict management styles. The efficacy for classroom management was measured by using the Emmer and Hickman's (1990) scale. The findings suggested that teacher efficacy

for classroom management was a strong significant predictor of the mutually focused conflict management styles (Morris-Rothschild & Brassard, 2006).

One issue related to the measure of teacher efficacy using the TSES, is that it measures general efficacy that is not task-specific or subject specific. For example, a math teacher may feel a high sense of efficacy in his or her ability to teach algebra and fractions to unmotivated students, but feel less efficacious when teaching geometry to gifted students (Vasquez, 2009). This general nature can be considered a benefit when researchers are interested in comparing teacher efficacy for teachers of different subjects or grades, and it can be considered a limitation when more situation-specific information is important.

International Validation Studies of the TSES

There have been several attempts made to examine the latent structure of the TSES for pre-service and in-service teachers in different countries. There is a frequent use of the TSES by international researchers.

Fives and Buehl (2010) administered the long form (24-items) of the TSES to 102 in-service and 270 pre-service teachers from the mid-Atlantic region of the United States to examine the factor structure of the TSES for both the short form and the long form. They found that the TSES as a three-factor structure is appropriate for in-service teachers while one factor accounts for the variability in the efficacy measure for pre-service teachers. Their results are consistent with the findings of Tschannen-Moran and Woolfolk Hoy (2001).

Fives and Buehl (2010) also found that means and reliability for both the short and long forms are similar to what was found in the Tschannen-Moran and Woolfolk Hoy (2001)

validation study. They suggested that either the short or long form can be used for pre-service or in-service teachers (Fives & Buehl, 2010).

Duffin, French, and Patrick (2012) conducted a study to validate the factor structure of the TSES with beginning pre-service teachers in the United States. They found that pre-service teachers' responses do not differentiate among the three aspects of efficacy measured by the TSES. Furthermore, they found high inter-factor correlations. Their study provides additional evidence of unidimensionality of the TSES for pre-service teachers, similar to Tschannen-Moran and Woolfolk Hoy (2001).

Klassen, Bong, Usher, Chong, Haun, Wong, and Georgiou (2009) tested the validity of the TSES in five countries: Canada, United States, Korea, Singapore, and Cyprus. They used the 12-item short form in their study. The TSES was translated into Greek for Cyprus teachers and Korean language for the Korean teachers while teachers from United States, Canada, and Singapore completed the form in English. There were a total of 1,212 elementary, middle, and secondary school teachers from five countries. The reliabilities of the composite scores of the TSES for Canada, United States, Korea, Singapore, and Cyprus were .83, .87, .92, .94, and .89 respectively. Subscale-wise reliabilities were also stable across the five countries. The results suggested that the factor structure of the TSES is not only valid for the teachers of the United States but also a valid efficacy instrument for Canadian, Korean, Singaporean, and Cyprus teachers. They found consistent results with Tschannen-Moran and Woolfolk Hoy (2001) regarding the three factors of the TSES (Klassen et. al., 2009).

O'Neill and Stephenson (2012) conducted a study on the TSES in Australia. A total of 573 pre-service primary teachers participated in the study. The participants were from various

Australian institutes that offered undergraduate primary teaching programs. The reliability for the 24-item long version was .94. The study found a higher efficacy mean of 6.95 for Australian pre-service teachers as compared to other international pre-service teachers (Charalambous et al., 2008; Fives & Buehl, 2010). The study did not find any significant associations between efficacy and gender for pre-school teachers. The factor loadings for the TSES were ranging from .67 to .81. There were no further results for the latent structure of the TSES reported.

Nie, Lau, and Liau (2012) attempted to examine the factor structure of the TSES in the context of Singapore. The data were collected from 109 full-time, in-service primary and secondary school teachers using the long 24-item version of the TSES. The authors used the English version of the TSES scale, however the items were rephrased to make them appropriate for a Singaporean population. A principal component analysis with an oblique rotation revealed a three-factor solution using the Kaiser's rule (Kaiser, 1960) of eigenvalues greater than 1. These three factors accounted for 75.66 % of the total variance (Nie, et. al., 2012). Based on subsequent EFA and CFA results, the revised version (Singapore context) of the TSES produced three factors. According to the authors, the 12-item revised version of the TSES had a good internal consistency; however the reliability coefficient was not reported in the paper. These 12-items were the same items adopted for the TSES short form, however the authors changed some words to better fit each item in a Singapore context. The inter-scale correlations were in the range of.60 to .68 among the three subscales.

Charalambous, Philippou, and Kyriakides (2008) conducted a study to validate the 24-item TSES for pre-service mathematics teachers in Cyprus. There were 89 pre-service teachers for elementary education (grade 4-6) who participated in the study. The reliability of the three

subscales: instructional strategies, classroom management, and student engagement was relatively higher (.96, .97, and .98 respectively). The factor analysis results produced a two-factor solution for pre-service mathematics teachers. Two items did not significantly load on any factor. The first factor (14 items) was labeled as instructional skills with factor loadings ranging from .52 to .78 while the second factor (8 items) was labeled as classroom management in mathematics with factor loadings ranging from .63 to .86. When authors reanalyzed the 22-item scale, it produced two-factor solution with 60% of the variance which further strengthen their result of two-factor solution for pre-service teachers. The authors concluded that preschool mathematics teachers have differing perceptions of their efficacy with instructional strategies and classroom management constructs. This finding contradicts the Tschannen-Moran and Woolfolk Hoy's (2001) argument that pre-service teachers are unable to discriminate among efficacy levels for the subscales of the TSES (Charalambous, Philippou, & Kyriakides, 2008).

Charalambous, Philippou, and Kyriakides's (2008) also administrated the TSES for the 89 preschool mathematics teachers two times; first, at the start of a teacher training program' and second, at the end of their program. As expected, the preschool teacher beliefs for instructional strategies was higher at the end of the teacher training course (M = 7.05, SD = 1.05) as compared to their belief for instructional strategies at the start of the program (M = 5.61, SD = 1.02). Similar results were found with the efficacy of classroom management that indicated a higher efficacy belief for classroom management at the end of the teacher training course (M = 7.01, SD = 1.25) compared to preschool teachers' belief for classroom management at the start of the program (M = 5.71, SD = 1.26). The authors did not test for statistical significance differences for these two repeated measures.

Teacher Sense of Efficacy and Student Achievement

Generally, teacher self-efficacy indicates teachers' beliefs in their ability to influence their students' achievement (Skaalvik & Skaalvik, 2010). Several studies have been conducted to investigate the relationship between the levels of teachers' sense of efficacy and student achievement (Holzberger, Philipp, & Kunter, 2013; Moore & Esselman, 1992; Ross, 1992).

One of the first attempts to investigate the relationship between teacher efficacy and student achievement was conducted by Armor et al. (1976). In their study of 356 black children in the U.S., they found that teacher sense of efficacy is a significant predictor of children's reading performance, $b = 0.13$, $t = 2.54$, $p < .001$. When including teacher sense of efficacy, classroom settings, program content, and implementation of strategies in the model, Armor et al. (1976) found that these predictors accounted for 70% of variance in the reading achievement of the black minority group. Authors did not provide the partial correlation between teacher efficacy and student achievement.

Armor et al. (1976) conjectured from their findings that teachers with a high sense of efficacy can influence student achievement in reading. Capra, Barbaranelli, Steca, and Malone (2006) also found a significant relationship between teachers' sense of efficacy and student achievement. Rose and Medway (1981) using the TLC instrument for 89 female fourth grade teachers found that teacher efficacy beliefs can influence teachers' behavior in the classroom, and teacher efficacy can predict their willingness to adopt and implement new instructional strategies in the classroom (Rose & Medway, 1981).

Tracz and Gibson's (1986) study used the TES (Gibson & Dembo, 1984) instrument with U.S. teachers, and they found there a significant correlation between the personal teacher efficacy of 4-6 grade teachers and student achievement. Anderson, Green and Loewen (1988) assessed the correlation between the TES (Gibson & Dembo, 1984) and Canadian Achievement tests. The participants were elementary school teachers of grade three and grade six. The results suggested that there is significant correlation between personal teacher efficacy and reading score of grade three but no correlation between personal teacher efficacy and reading score of grade six. The author did not provide the correlation coefficients. Ashton and Webb's (1986) study proposed that teacher efficacy is related to instructional practices and student performance. They found a significant positive relationship between students' mathematics scores and teachers' sense of efficacy. Adu, Tadu, and Eze's (2012) study in Southwestern Nigeria also found teachers' self-efficacy is significantly correlated with secondary school student performance, $r = .38$, $p < .001$.

Khan (2011) evaluated the relationship between teachers' efficacy and students' achievement at the secondary level in Pakistan. She found that teachers with higher efficacy tend to have higher expectations of students which are also significantly related to higher student achievement. The author did not provide statistical values in her paper. Akram and Ghazanfar (2014) found relationship between teacher efficacy and students' GPA (grade points average). Their results with 193 responses revealed that there is strong positive relationship between self-efficacy and students' GPA, $r = .76$, $p < .001$.

Khan (2012) used the Urdu translated scale of TSES in the Attock district of Pakistan. He found significant and strong correlations in teachers' self-efficacy and student achievement in

both math ($r = .71$, $p = .002$) and reading scores ($r = .91$, $p < .001$). Because of the strong link between the perceived efficacy of teachers and student achievement, mechanisms for increasing teacher efficacy beliefs should be investigated in the search for practical strategies that might increase student performance.

Alrefaei (2015) examined the relationship between teacher sense of efficacy and student achievement in fifth grade mathematics and science subjects using the short form TSES for 62 fifth grade teachers in Arkansas, USA. The findings indicated that there was no significant relationship between teachers' sense of efficacy and fifth grade students' mathematics scores on Arkansas benchmark tests, $r(29) = .33$, $p = .07$. There was also no significant relationship between teachers' sense of efficacy and fifth grade students' science scores on the Arkansas benchmark tests, $r(29) = .33$, $p = .07$. It should be noted that the power was relatively small given the limited sample size.

Teacher Sense of Efficacy and Job Satisfaction

Teacher self-efficacy is also related to teacher job satisfaction. Klassen and Chiu's (2010) study included 1,430 practicing teachers teaching at elementary and high schools in western Canada. They concluded that teachers with high levels of self-efficacy tend to have high levels of job satisfaction ($r = .69$, $p < .001$). A significant strong relationship ($r = .78$, $p < .001$) between the level of teacher efficacy and students' attitudes toward school was also found in a study of 679 male and female teachers from 23 primary and junior schools from the capital of Jordan (Al-Alwan & Mahasneh, 2014). They also found that 56% of the total variance in student attitudes' towards school can be explained by teachers' self-efficacy ($F = 6.12$, $p < .001$). Teachers with high efficacy also tend to report being more committed to their teaching (Coladarci, 1992).

Teacher self-efficacy also serves as a predictor for job performance. Ereño and Nunez (2014) conducted a study in Phillippe to examine the relationship between teacher's sense of efficacy and their job performance in higher education institutes. Teacher sense of efficacy was measured by the TSES (Tschannen-Moran & Woolfolk Hoy, 2001) and performance of the faculty members was obtained from the student evaluation reports of the teachers for two semesters. The findings revealed that teachers who were highly efficacious performed better or had higher student evaluations in higher education institutes than teachers with low self-efficacy.

Teacher Sense of Efficacy and Teacher Demographic Characteristics

Teachers' demographic characteristics such as gender, age, teaching experience, and teaching grade level have also been linked to teacher self-efficacy levels. Research indicates there is variation in teachers' levels of efficacy and their gender. When reviewing the literature, the level of efficacy is reportedly higher in female teachers than in male teachers. Edwards, Green and Lyons's (1996) study concluded that male teacher (N = 43) have lower self-efficacy than female teachers (N = 379) in general, $F(1, 420) = 5.91$, $p = .02$. However, caution should be taken when interpreting their results due to the large difference in the size of the male and female groups, and more specifically the small number of male teachers. Penrose, Perry and Ball's (2007) study found no significance difference between female government teachers (N = 135, M = 72.33, SD = 10.08) and male teachers (N = 75, M = 70.34, SD = 9.90), $t(207) = 1.38$, $p = .17$ in Victoria, Canada.

Teacher qualifications is also related to teacher's self-efficacy. A study administering the TSES to 62 science and math teachers from Arkansas (U.S.) found that teachers who have a bachelor's degree (M = 7.84, SD = 0.49) tend to have a higher efficacy level than teachers who

have a masters' degree (M = 7.47, SD = .74), t(57.8) = 2.29, p = .025, d = 0.58 (Alrefaei, 2015). The reason for lower efficacy with a higher degree needs to be studied further to understand the reason for the relationship.

Research indicates teachers who teach lower grade levels have a higher sense of efficacy than teachers at the secondary level. For example, elementary teachers tend to have higher level of self-efficacy than senior high teachers, F(2, 418) = 5.42, p < .005 (Edwards, Green, & Lyons, 1996).

Hunt-Ruiz (2011) study on the early stages of efficacy indicated that pre-service teachers reported high levels of teaching efficacy, while teachers' level of efficacy declined as their years of teaching experience increased. Similar findings were reported by Klassen and Chiu (2010) that more experienced teachers tend to have lower levels of teaching efficacy. However, Alrefaei (2015) found no significant difference in teacher's sense of efficacy and their teaching experience.

Teachers with a higher sense of efficacy encourage their students and motivate them to come to school and learn. Students' attitudes toward schools is significantly related to teachers' self-efficacy, r = 78, p < .001 (Al-Alwan & Mahasneh, 2014). Teachers with higher efficacy in regard to student engagement may be more capable of engaging with students, thus it positivity impacts students' attitudes towards attending school.

Teacher self-efficacy is also associated with teacher burnout. Several studies have been conducted to investigate the relationship between the level of teacher efficacy and teacher burnout. Khezerlou (2013) found that teacher self-efficacy is a moderate predictor of job

burnout. Skaalvik and Skaalvik (2010) evaluated the relationship between teacher efficacy and teacher burnout among 246 elementary school teachers in Norway. They found teacher self-efficacy is a significant predictor of teacher burn-out with regression coefficients in the range between -.32 and -.40.

Relationship with Teachers' Demographics in Pakistan's Context

In Pakistani's context several researchers have compared teacher sense of efficacy and teacher demographic characteristics. Shaukat and Iqbal (2012) used the TSES instrument to compare subscale-wise mean differences on teachers' demographic characteristics. They found a significant difference in *classroom management* between male teachers (M = 27.54, SD = 4.37) and female teachers (M = 26.28, SD = 3.91), $t(196) = 2.11$, p =.03 but did not find significant differences in *student engagement* and *instructional strategies* between male and female teachers. The study did not find any difference in the efficacy of *student engagement* and *instructional strategies* between elementary and secondary school teachers however there was a significant difference between elementary (M = 27.67, SD = 3.82) and secondary school teacher (M = 26.15, SD = 4.50) in the efficacy of *classroom management*, $t(196) = 2.56$, p =.01 (Shaukat & Iqbal, 2012). Ahmad, Khan, and Rehman's (2015) study found a significant difference between male and female teachers at the elementary level, $t(60) = 3.14$, p = .002 using the TSES composite scores. Elementary school female teachers (M = 79.12) reported a higher mean on the general teaching efficacy level than elementary male school teachers (M = 75.23). Haq and Akhtar (2013) attempted to evaluate the teacher sense of efficacy with respect to school level and teaching experiences. They collected responses from 818 school teachers from the Punjab province (high school = 348, middle level = 307, primary = 163). They used the long form (24-

item) of the TSES with the Urdu translated version. The results showed a significant difference in teacher sense of efficacy among high (M = 139.65, SD = 33.38), middle (M = 137.32, SD = 25.94), and primary (M = 146.10, SD = 25.97) school teachers, $F(2, 815) = 4.67$, $p = .01$. Similar to other studies, the primary teachers reported the highest efficacy levels. Note that they used the Urdu translated version without a validity assessment of the instrument.

Butt, Khan, and Jehan (2012) evaluated the relationship between English teachers' self-efficacy and students' performance based on the gender of teachers for 10th grade government school teachers in the Khyber Pakhtunkhwa province of Pakistan. They administered the Urdu translated version of the TSES and students' achievement test. Authors did not provide the information about the type of achievements they used in their study. The results indicated that the female English teachers' sense of efficacy was greater than the efficacy level of male teachers. In addition, male and females English teachers with a high sense of efficacy believed they had the ability to impact students' motivation, whereas teachers with low efficacy reported a low ability to influence their students (Butt, Khan, & Jehan, 2012).

The components of teacher self-efficacy are important to study and understand in regard to teacher development and their effectiveness in the classroom. The TSES, as one of the more widely used teacher self-efficacy instruments, is beginning to be used in Pakistan even though no validation study has been conducted with this population on its appropriateness. The purpose of this study is to evaluate the use of the TSES and assess its factor structure with pre-service and in-service teachers in Pakistan.

CHAPTER III

Research Methodology

This study included a validation assessment and provided an evaluation of the latent structure of the Teacher Sense of Efficacy Scale for in-service teachers and pre-service teachers in Pakistan. Four different models were tested based on prior studies found with in-service and pre-service teachers from other countries (i.e., US, Australia, Malaysia, Turkey, and Norway). This chapter presents research questions, participants' details, sample size, description of instruments, research design, data collection, and data analysis techniques.

Participants

Participation in this study was on a voluntary basis and was open for male and female: 1) in-service teachers, who are teaching various courses at elementary and secondary levels in Pakistani public schools for at-least one year, and 2) pre-service teachers enrolled in education-based programs for elementary and secondary training (i.e., B.Ed., M.Ed.) at the university who have not officially taught within a public or private school. However, pre-service teachers who participated in this study have completed their 40 hours of internship in schools.

Data were collected from the four provinces of Pakistan including Islamabad capital territory and Azad Kashmir. There were a total of 557 in-service teachers' responses including 31% male 69% female in-service participants. There were total 433 pre-service teachers including 27% male and 73% female pre-service teachers participated in this study. Complete details about the participants is provided in the chapter IV.

Sample Size. Sufficient sample size is an important factor for many statistical procedures when conducting a validation study, and specifically an influential factor in a Structure Equation Modeling (SEM) framework where **"**the adequacy of the test statistics is likely to be influenced by sample size" (Hoyle, 1995, p.87). Sample size is a vital factor in evaluating the appropriateness of the SEM model. It determines the extent to which the model evaluation can be trusted (Hoyle, 1995). Generally, there are two approaches to determine the required sample size for the latent model analysis: minimum absolute sample size and participants to variable ratio. In the case of SEM, researchers usually recommend that the larger the sample size, the better. There is no concise agreement to determine one appropriate sample size. Suggested minimum sample sizes include from 3 to 20 participants per observed variable or entire sample size ranges from 100 to 1000 (Mundfrom, Shaw & Tian, 2005).

Bentler and Chou (1987) proposed a minimum of five subjects per free parameter under normally distributed data. However, they recommended the ideal ratio is 10 participants or even 15 participants per variable for better estimation. A similar sample size is proposed by Stevens (2002). Mueller (1997) also recommended a 15:1 ratio (15 participants per item) or at least a 10:1 ratio.

Baldwin (1989) and Lomax (1989) recommended that an overall sample size should be at least 200. Kunnan (1998) concluded that a sample size less than 150 may produce unstable estimation and threaten external validity. A sample size of at least 400 is recommended for robust maximum likelihood estimation (Boomsma & Hoogland, 2001). Schumacker and Lomax (2010) found sample sizes in between 200 to 500 in most SEM studies. When reviewing the

literature about sample size determination for the SEM, the most common rule of thumb is either (a) minimum 200 participants or (b) at least a 10:1 or 15:1 participant ratio per item.

For this study, a minimum of 15 participants per item was selected for stable estimation. Therefore, the overall targeted sample for this study was around 1000. Given that the latent model analyses were conducted for both groups of pre-service and in-service teachers, a 500 sample size for each group was sufficient for reliable and stable SEM statistics estimations. The sample size meets the minimum sample size requirement given by the researchers listed with approximately 15 respondents per item for the 24-item long form of the TSES. The complete information about sample size for in-service and pre-service teachers is discussed in chapter IV.

Sampling Method. Although the sampling method was convenience due to the volunteer nature of the surveying, an attempt was made to obtain data from a diverse representation of Pakistani in-service and pre-service teachers and background information was reported to describe those participating. According to Pakistan Education Statistics (2011), the education system of Pakistan is comprised of 270,825 (72% public and 28% private) institutions with the help of 1,507,100 teachers with more female (55%) teachers than male (45%) teachers in education. There are 184 teacher training institutes in Pakistan, including 82% public and 18% private institutes (AEPAM, 2011). In regard to teacher gender, 34% of students enrolled in teacher training institutes are female, whereas 66% are male students (AEPAM, 2011). This figure indicates that female prospective teachers may appear to be more committed to joining the teaching profession than male prospective teachers.

The survey participants were from major cities representing the provinces of Pakistan including Sindh (Karachi, Hyderabad, Jamshoro, and Khairpur), Baluchistan (Queeta, Pisni, and

Kalat), Punjab (Lahore, Mianwali, Naroval, and Rawalpindi), Islamabad Capital Territory (Islamabad city and Federal Area), Khyber Pakhtunkhwa (Peshawar), and Azad Kashmir. The following table shows area, population and number of teachers (public, semi-public, and private sector) at each level by province.

Table 2

Number of Teachers by Province and School Level (Source: AEPAM, 2011)

Province	Area (KM^2)	Population [a]	Primary	Middle	High	Higher Sec. [b]	Total
Punjab	205,344	96,545,293	50,008	93,510	125,494	22,051	291,063
Sindh	140,914	42,187,865	51,886	32,944	72,410	12,013	169,253
Khyber Pakhtunkhwa	74,521	23,680,359	13,225	9,331	15,014	5,474	43,044
Baluchistan	347,190	13,160,000	6,246	4,288	11,037	666	22,237
AJ &K	11,639	3,676,426	888	2,223	4,190	624	7,925
GB	72,520	1,228,650	608	631	905	98	2,242
ICT	906	1,420,983	1,923	586	1,871	1,804	6,184
		Total	124,784	143,513	230,921	42,730	541,948

[a] National Institute of Population Studies (NIPS), Government of Pakistan Projection (2011).
[b] Higher Sec. refers to higher secondary teachers of grade 11 and 12.

Instruments

Participant's demographic questionnaire. Participants were also asked to complete a demographic questionnaire. Demographic data were used for descriptive and inferential statistics such as MANOVA for mean comparisons of teachers' sense of efficacy based on teachers' demographics. The demographic questionnaire (see appendix A) provides information regarding teachers' personal characteristics such as gender, age, teaching grade level, teaching experience, and academic qualification.

Teacher Sense of Efficacy Scale (TSES; Tschannen-Moran & Woolfolk Hoy, 2001). A current instrument used to measure teacher efficacy level was developed by Tschannen-Moran and Woolfolk Hoy (2001), called 'Teacher Self-Efficacy Scale (TSES)' also known as the 'Ohio State Teacher Efficacy Scale (OSTES)'. There are two versions of the TSES scales: a short version consisting of 12 items, and a long version consisting of 24 items. The long-form TSES was used in this study which also comprised of the items on the short-form TSES.

The 24 items of the TSES (see appendix B) provides information on the efficacy levels of three sub-scales (i.e., classroom management, instructional practices, and student engagement). Each sub-scale in the long-form TSES contains eight items. Efficacy on *Student Engagement* was measured by item 1, item 2, item 4, item 6, item 9, item 12, item 14, and item 22; efficacy on *Instructional Strategies* subscale was measured by item 7, item 10, item 11, item 17, item 18, item 20, item 23, and item 24; while efficacy on *Classroom Management* was measured by item 3, item 5, item 8, item 13, item 15, item 16, item 19, and item 21.

In each item of the TSES, respondents are asked about the degree to which they have an impact on certain instructional or student outcomes. Each item is formatted on a 9-point Likert type scale ranging from 1 (nothing) to 3 (very little) to 5 (some influence) to 7 (quite a bit) to 9 (a great deal). Formal permissions have been obtained from both authors Tschannen-Moran (appendix H) and Woolfolk Hoy (appendix I) to use the TSES instrument in this study.

Behavior Management Strategies Scale (BMS; Nie, Lau & Liau, 2012). Nie, Lau and Liau (2012) refined the behavior management strategies scale adopted from the Mathematics Enhancement Classroom Observation Record Scale (MECORS; Schaffer et al, 1998). The BMS is used to measure teachers' strategies to cope with classroom behavior. This scale contains 7

items with a 5-point Likert format from 1 (*never*) to 5 (*always*). A sample item from the BMS is *"I establish specific rules and consequences for student misbehavior"* (Nie, Lau & Liau, 2012, p.421). A study conducted by Nie, Lau and Liau (2012) found that there is a significant, moderate correlation between the BMS and the efficacy in *Classroom Management* (CM) in teachers from Singapore, $r = .52$, $p < .001$. Formal permissions have been obtained from authors to use the instrument in this study (see appendix J).

Personal Teaching Efficacy (PTE; Gibson & Dembo, 1984). *Personal Teaching Efficacy* (PTE) from Gibson and Dembo's (1984) the Teacher Efficacy Scale (TSE) appears to measure a construct similar to *Instructional Strategies* (IS) of the TSES (Tschannen-Moran & Woolfolk Hoy (2001). The PTE is a 9-item scale with a 6-point Likert format from 1 (*Strong Disagree*) to 6 (*Strongly Agree*). The items of the PTE measure teachers' efficacy level on instructional strategies and personal engagement with unmotivated and/or underperforming students. A sample item of the PTE is "When a student is having difficulty with an assignment, I am usually able to adjust it to his/her level" (Gibson & Dembo, 1984, p. 581). There is one item that appears to measure efficacy on student engagement (i.e., *When I really try, I can get through to most difficult students*) and one item that seems to measure teacher efficacy on classroom management (i.e., *If a student in my class becomes disruptive and noisy, I feel assured that I know some techniques to redirect him quickly*), however this scale seems to most closely align with teacher sense of efficacy on instructional strategies with seven of the nine items measuring instructional strategies components of teaching. Given that no operation definition is provided for the PTE scale, a personal review of the content contained in the items was used. A previous study shows there is a moderate correlation between the PTE and *Instructional Strategies* (IS), $r = .62$, $p < .001$ (Tschannen-Moran & Woolfolk Hoy, 2001).

Instructional Management Scale (Martin & Sass, 2010). No appropriate alternative standardized scale that measures teacher self-efficacy on student engagement was found. The two complicating factors were identifying a scale with a similar operational definition (based on the internship of the items) in addition to a scale that asked for the teachers' perceptions rather than an observation-based instrument on student behavior. After an extensive literature review, the *Instructional Management* scale (Martin & Sass, 2010) was selected for testing the hypothesized relationships between it and the three subscales of the TSES. Based on the item content, it is hypothesized that Martin and Sass's *Instructional Management* scale should be most correlated with the TSES *Student Engagement* scale, followed by the TSES *Instructional Strategies* scale. It should be least correlated with the *Classroom Management* scale from the TSES, although not uncorrelated. Martin and Sass's (2010) *Instructional Management* (IM) scale contains six items on a 6-point Likert scale from *Strong Disagree* to *Strongly Agree*. Martin and Sass (2010) found that the correlation between their *Instructional Management* scale and efficacy in *Student Engagement* (TSES) scale was moderate for a sample of 550 certified teachers from the United States, $r = .65$, $p = .005$. The *Instructional Management* scale was also found to be significantly correlated to the *Instructional Strategies* ($r = .59$, $p = .005$) and *Classroom Management* ($r = .51$, $p = .005$) subscales of the TSES.

Data Collection Procedure

Once official approval was received from the University of Arkansas Institutional Review Board (IRB) for human subject research (appendix F), data collection from the in-service and pre-service teachers began. The data collection process took three months during the fall of 2015. In this period of three months, the researcher visited universities that offer teacher training

degree programs (B.Ed. / M.Ed.) in Pakistan. There were two formats used for data collection: paper based and online versions of a survey. Paper printed survey forms were distributed to the participants in their respective institutes while the online version of the survey was used for institutes in remote areas. The online-version of the survey was administrated using Qualtrics online survey software (Qualtrics.com). A recent study conducted by Ravert, Gomez-Scott, and Donnellan (2015) compared the psychometric properties of paper-based and computer survey data. They found minor differences in the homogeneity of the paper-based and computer survey data. They did not find any evidence of differences between these two methods in terms of acceptance rate, proportion of missing data, or internal consistencies (Ravert et al., 2015).

During the paper-based data collection process, the researcher explained the purpose and the significance of the study to the participants before asking for their voluntary agreement to participate in this study. The instrument was self-administered after the researcher distributed it among participants (teachers). The questionnaire took around 20 minutes to complete. Statements in the instrument were self-explanatory; however the researcher was there to clarify if participants had questions. The researcher made sure that participants had a comfortable and pleasant environment before completing the questionnaire. The online source was the second source of data collection. The online link for the survey was emailed to teachers to get sample participation from remote parts of Pakistan. Participants were invited to voluntarily participate in the study. The survey link remained active for one month from the date of invitation. A reminder was sent to potential participants after two weeks from the original invitation. After three months of data collection using both the online and paper based formats, the data were aggregated.

Data Analysis Procedures

The data for many of the statistical procedures were analyzed using SAS 9.4 (SAS Institute; Cary, NC) software, but the structure equation modeling analyses were conducted using EQS 6.2. A range of statistical tests was conducted to investigate the validity of the TSES scales for Pakistani in-service and pre-service teachers. Crocker and Algina (1986) define validity as the process by which a researcher collects evidence to support the types of inferences that are to be drawn. First, items underwent a review by content experts within the Pakistani culture to evaluate their appropriateness for measuring the stated objectives with a focus on whether the items would be interpreted appropriately by teachers from Pakistan. Next, data were collected and evaluated for errors. Analyses were then conducted to assess reliability and proceeded with the measures of validity including the analysis of latent factor structures and convergent and discriminant validity comparisons. Last, multivariate analyses of variance (MANOVA) were conducted to examine the differences in efficacy beliefs based on teachers' characteristics such as gender, teaching grade level, and qualifications. The overall *F* test statistical significance level was set at .05 for each MANOVA. The analyses for reliability and validity of the instrument for both pre-service and in-service teachers were divided into the following four sections.

Content validation procedure. Prior to collecting data from participants, an independent evaluation of the TSES items was conducted by seven education professionals and content experts who have educational experiences in Pakistan and international institutes. The purpose of this expert judgment was to investigate item validity from a cultural prospective in addition to item construct match. Item validity evaluates the degree to which an item measures what it is

hypothesized to measure. The focus of this study was to investigate the use of the three scales on the TSES for in-service and pre-service teachers in Pakistan.

Content validity is the degree to which an item or set of items adequately and sufficiently measures skills or behaviors which it is defined to measure. An instrument is considered to have evidence to support content validity if its items constitute language, structures, and skills to measure a particular construct that also fits the targeted population (Brown & Hudson, 2002; Hughes, 2003). A recommended method to establish content validity is to have experts review items and to determine each item's relevance to its anticipated objective (DeVellis, 2003; Haynes et al., 1995). The content validation process incorporates both qualitative and quantitative methods (DeVellis, 2003). Following the mixed-method recommendations for content validity, each item of the TSES was examined qualitatively through focus group discussion and quantitatively through a three-point scale using index of item objective congruence (Rovinelli & Hambleton, 1977). Content experts were asked to evaluate the items on the degree to which they measure the hypothesized scales, but to also use their familiarity with the Pakistani culture to identify if any items may be interpreted differently than intended due to different cultural interpretations of certain words / phrases or differences in the Pakistani educational context.

Content validity can be assessed for unidimensional items and also for multidimensional items (Turner & Carlson, 2003). A unidimensional item is an item which measures either one construct or a set of constructs that are essentially unidimensional while a multidimensional item measures more than one construct and respondents vary in their responses to those constructs. In the current study, all 24 items of the TSES are designed to be unidimensional in that each item measures only one construct out of three proposed constructs.

The Index of Item Objective Congruence (Rovinelli & Hambleton, 1977) procedure was used for assessing item validity using experts' ratings. The Index of Item Objective Congruence (IIOC) is a blind rating of the degree to which an item measures each construct. This process was a blind-review which means that the experts were not told which item relates to which constructs. The content experts evaluated each item as it corresponded to three listed objectives (i.e., *Student Engagement*, *Instructional Strategies*, and *Classroom Management*) by giving it a rating using a three-point scale: +1 if an item clearly measures the listed objective, -1 if an item does not clearly measure the listed objective, and 0 if an item is an unclear measure of the listed objectives. The IIOC for unidimensional items was calculated by the following simplified version of the formula provided by Crocker and Algina (1986, p. 221),

$$I_{ik} = \frac{N}{2N-2}(\mu_k - \mu)$$

where I_{ik} is the index of item-objective congruence for item *i* measuring the kth objective, N = the number of objectives; μ_k is the experts' mean rating of item *i* on the kth objective; and μ is the experts' mean rating of item *i* on all objectives.

The highest item-objective congruence value is 1.00 when an item is clearly matched to one and only one construct. The lowest IIOC value is -1.00 which indicates that an item is clearly not a measure of the hypothesized construct and a clear measure of all invalid constructs. This information was used to aid in the interpretation of any items not fitting into the hypothesized latent models for in-service and pre-service teachers in Pakistan.

Reliability analysis. The Cronbach's (1951) alpha internal consistency reliability analysis procedure was used for the full sample, and then separately for in-service teachers and

pre-service teachers. The reliability coefficient indicates the consistency of the instrument. As a general rule, internal consistency of ≥ .80 is considered an acceptable reliability in social sciences. The Cronbach's alpha reliability coefficient is defined as (Crocker & Algina, 1986)

$$\rho_{\alpha} = \left[\frac{k}{k-1}\right]\left[1 - \frac{\sum \sigma_i^2}{\sigma_x^2}\right]$$

Where k = number of items, σ_i^2 = variance for item i, and σ_x^2 = variance of total test scores.

Construct validity. Construct validity was evaluated by confirmatory factory analysis in the SEM framework. A Confirmatory Factor Analysis (CFA) shows the relationship between indicators and latent traits. The purpose of CFA is to evaluate whether hypothesized constructs among a set of variables are supported by the observed data. CFA was used to validate the hypothesized factor structure of the TSES scale and to compare it to factor structures found with other populations. The CFA procedure was used to test following models:

1. Model I (Three-factor model for in-service teachers - long-form TSES)
2. Model II (Three -factor model for in-service teachers - short-form TSES)
3. Model III (One-factor model for pre-service teachers - long-form TSES)
4. Model IV (One-factor model for pre-service teachers - short-form TSES)

When researchers use the sample variance-covariance model, the chi-square ($\chi 2$) test is used to evaluate the model fit that the population covariance and model-implied covariance are equal (Gerbing & Anderson, 1993).

$$\mathrm{H_o}: \sum \text{population} = \sum \text{model}$$

However, the chi-square test is sensitive to sample size. Therefore, a small difference with a large sample size between the sample and model implied covariance may lead to model rejection (Gerbing & Anderson, 1993; Millsap, 2007). When the model is rejected, it is concluded that the model does not fit the data well. Thus, these sample size and power issues lead researchers to formulate alternative fit indices that control the effect of large sample size (Bentler & Bonett, 1980; Jöreskog & Sörbom, 1981).

This research evaluated each model fit by using a multi-index strategy as recommended by Byrne (1994) and Hu and Bentler (1999). Several validation studies have been conducted using the multi-index strategy (e.g., Duffin, French, & Patrick, 2012). The model fit indices are compared in terms of the degree of acceptance rather than a yes or no conclusion. The purpose of evaluating data model fit is to characterize strengths and weaknesses of the model with respect to alternative hypotheses or competing models. The fit indices are generally categorized into three groups: absolute fit indices, parsimonious fit indices, and incremental fit indices.

Absolute fit index values evaluate the overall inconsistency between the observed and hypothesized models. The absolute fit indices include:

1. Model *T* statistics or (chi square tests) associated with sample size. The smaller *T* statistics provide better model fit. Results are stable with large sample sizes.
2. Standardized Root Mean Squared Residual (SRMR) is a function of standardized residuals, a difference between the observed and expected matrix. The small value indicates better model fit.
3. Goodness-of-Fit Index (GFI) is designed as a coefficient of determination *R* squared, the larger GFI provides better model fit. The GFI value ranges between 0 and 1. The GFI is

one of the most well-known indices (Jöreskog & Sörbom, 1981). The GFI is more appropriate than Normed Fit Index (NFI) in finite samples (Tanaka & Huba, 1989).

The parsimonious model fit indices provide improvements in model fit as more parameters are added. The parsimonious (aks Improving Absolut Fit) indices include:

4. Root Mean Squared Error of Approximation (RMSEA) provides mean square error of approximation. The smaller value indicates better model fit. It also provides confidence intervals.
5. Akaike Information Criterion (AIC) is used for non-nested models. The AIC is a good model fit index to compare for non-nested models, however it penalizes with the use of too many parameters (Akaike, 1987). A relatively smaller AIC value suggests better model fit.

Incremental Fit Indices provide evidence of a model's absolute or parsimonious fit values relative to a baseline (null) model where all observed measurement variables are uncorrelated (Hu & Bentler, 1999). The larger incremental fit indices provide better model fit. The incremental fit indices include:

6. Comparative Fit Index (CFI) is a widely used index among the other indices used in structural equation modeling (SEM). This study will use the criteria provided by Hu and Bentler (1999)'s recommendations for combined indices rule SRMR $\leq$ .09, NFI $\geq$ .90, and CFI $\geq$ .95 based on their empirical work.

The Hu and Bentler's (1999) criteria for good model fit indices are provided in the table 3. These criteria were used to determine if a model is to be rejected or accepted.

Table 3

Hu and Bentler's (1999) Fit Indices Criteria to Retain a Model

Fit Indices	Criteria for 'acceptable' model fit	Criteria for 'good' model fit
NFI	.87	≥ .90
NNFI	.90	≥ .95
CFI	.90	≥ .95
GFI		≥ .90
SRMR		≤ .08
AGFI		≥ .90
RMSEA	≤ .08	≤ .06
Joint Criteria	NNFI, CFI ≥ .96 and SRMR ≤ .09	
	SRMR ≤ .09 and RMSEA ≤ .06	

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

Convergent validity. Further evaluation of the construct validity and convergent validity was assessed using the Multitrait-Multimethod (MTMM) procedure. The multitrait multimethod matrix is a two dimensional cross classification of traits and methods (Maas, Lensvelt-Mulders, & Hox, 2009) developed by Campbell and Fiske in 1959. The MTMM matrix is a correlation matrix that provides evidence of construct validity on a set of scales in the study (Campbell & Fiske, 1959). The multitrait multimethod approach has proven to be a significant process in the past half century of psychological research of differentiating between related scales' constructs (Byrne, 1994). In the current study, the MTMM matrix was used to compare the constructs being measured between the TSES and alternative measures of subscales hypothesized to be measuring the same or similar constructs. In their study, Tschannen-Moran and Woolfolk Hoy (2001) did not test the convergent and discriminant validity of the three subscales of TSES to provide

further empirical evidence for the appropriateness of the TSES measure (Nie, Lau, & Liau, 2012).

The MTMM matrix is also helpful to assess divergent validity. Divergent validity is established when the correlation between two different constructs is smaller than the relationship between two scales hypothesized to measure the same or similarly defined constructs. Since the Teacher Sense of Efficacy has three subscales that have moderate relationships among the scales, it can be difficult to identify alternative measures that are clear convergent and divergent validity comparisons for the original set of scales. Thus an a priori prediction of comparative relationship between the original and alternative sets of scales is important to the process. Moreover, to evaluate the correlations among the constructs without the impact of measurement error or lack of reliability in the scales, a second set of correlations corrected for attenuation was calculated. The correction for attenuation (aka: disattenuation) addresses the possible concerns about the impact of unreliability and the effect of measurement error on raw score correlations (Jensen, 1998). The disattenuated estimation between two scales can be calculated by following formula

$$\rho = \frac{Corr(XY)}{\sqrt{r_{XX} r_{YY}}}$$

where Corr(XY) is the correlation between two scales, r_{XX} is the reliability coefficient of first scale, and r_{YY} is the reliability coefficient of second scale.

The disattenuated correlations can be considered as a vital element in the validation process. The MTMM procedure in the current study added information for comparing convergent validity and discriminant validity coefficients for scales that have been identified as close to a match for the TSES subscales as could be determined by the current researcher given

that there are no published operational definitions for the TSES scales. Scale item groupings were used to identify the underlying latent constructs being measured, and these themes were used to identify alternative scales that appear to match the TSES in theoretical construct. The subscales selected for convergent validity comparisons were *Personal Teaching Efficacy* (Gibson & Dembo, 1984), *Behavior Management Strategies Scale* (Nie, Lau & Liau, 2012), and *Instructional Management Scale* (Martin & Sass, 2010).

CHAPTER IV

Data Analysis

This chapter presents sample size, response rates, and demographic information regarding pre-service and in-service teachers, who participated in this study. Data were analyzed using SAS 9.4 while the structure equation modeling procedures were tested in EQS 6.2 (see appendices K to L). Item-level, subscale-level, and overall scale level descriptive statistics for pre-service and in-service teachers are presented in this section. Results begin with IIOC analyses. The reliabilities of instruments were assessed by Cronbach's alpha, then confirmatory factor analyses are presented for the TSES short-form and the long-form. CFA analyses were conducted to test initial hypotheses, with a post-hoc EFA (Exploratory Factor Analysis) procedure used to investigate the number of factors for pre-service teachers. A multitrait multimethod (MTMM) matrix was used to assess the convergent and divergent validity of the TSES. Last, a comparison of sample subgroups was conducted using MANOVA procedures to test the mean difference in the efficacies measured by the three subscales for gender, age group, teaching grade-level, and teachers' qualification.

Survey Distribution and the Response Rate

The data collection process was conducted by the researcher. The survey forms (appendices A to E) contain the participant's demographic questionnaire, *Teacher Sense of Efficacy Scale* (Tschannen-Moran & Woolfolk Hoy, 2001), *Personal Teaching Efficacy* (Gibson & Dembo, 1984), *Behavior Management Strategies Scale* (Nie, Lau & Liau, 2012), and *Instructional Management Scale* (Martin & Sass, 2010). Formal permissions were taken from

original authors to use above mentioned scales in this study. After official approval from the University of Arkansas Institutional Review Board (IRB) for human subject research (see appendix F), data collection process started. Informed consent (see appendix G) requirement was fulfilled to ensure that participant had understood the nature of the study and voluntarily agreement to participant in this study. The researcher also took permission from the head of institutes before obtaining responses from potential participants.

After distribution of survey forms, the researcher explained the procedure for completing survey forms and was available to answer any participant questions. Most of the forms were collected on the same day while some of the filled survey forms were collected from in-service teachers after a week. There were 990 respondents completing the surveys which included 930 paper-based surveys (94%) and 60 online respondents (6%) included in the final aggregated responses. Table 4 shows the number of distributed paper surveys and response rate from in-service and pre-service teachers.

Table 4

Distributed Survey Forms and Response Rate from In-service and Pre-service Teachers

Province	In-service Teachers			Pre-service Teachers		
	Surveys Distributed	Survey Returned	Reponses Rate	Surveys Distributed	Survey Returned	Reponses Rate
Sindh	150	143	95%	200	194	97%
Punjab	130	116	89%	100	94	94%
Baluchistan	100	92	92%	60	58	97%
Khyber Pakhtunkhwa	65	59	91%	55	52	95%
Islamabad Capital Territory	68	65	96%	33	31	94%
Azad Kashmir (AJK)	28	26	93%			

Total	541	501	93%	448	429	96%

In an attempt to minimize coverage error and sampling bias (Dillman, Smyth, & Christian, 2009), participants' responses were collected from all four provinces of Pakistan including the capital of Pakistan and Azad Kashmir with a focus on high response rate for the surveys administered by the researcher. The response rate for data collection was over 93% for the paper-based surveys. The response rate from pre-service teachers was just slightly higher than the response rate from in-service teachers, however both would be considered high response rates for volunteer surveys. Figure 3 depicts the regions of Pakistan along with a visual representation of the number of participants who participated in this study. The size of circle in figure 3 represents the number of teachers who participated in this study. No data are obtainable for the response rate of the 60 online surveys.

There were a total of 557 in-service teachers' responses (including both online and paper-based method). Eight forms were discarded during data cleaning process. These forms were removed due to unclear responses or most of the fields were vacant. Thus, 549 in-service teachers' responses were used for data analysis procedures. There were a total of 433 pre-service teachers' responses. Ten forms were discarded due to missing values in the majority of items. Therefore, a total of 423 pre-service teachers' responses were used for data analysis.

Teacher Demographics Information

In-service teachers. There were 171 (31%) male and 378 (69%) female in-service teachers in this study. Teachers were from all major cities of Pakistan and from different grade levels. There were 448 (82%) permanent teachers and 101 (18%) temporary teachers in the service. The complete demographic information for in-service teachers is presented in Table 5.

Teaching experience of the in-service teachers varied from two years to 36 years with the average of 11.80 years and standard deviation of 8.49. In-service teachers' age were spread across all six age groups. For example, 20% in-service teachers' age were in the range between 25 and 29; 20% in-service teachers' age were in the range between 30 and 34; and approximately 19% in-service teachers' age were above 44 years. Two third of the in-service teachers had a masters' degree (N = 366) and about 26% of in-service teachers had a bachelor degree (see Table 5).

Pre-service teachers. There were 116 (27%) male pre-service teachers and 307 (73%) female pre-service teachers. Table 6 provides the demographics of pre-service teachers. Most of the pre-service teachers (75%) were less than 25 years old. While 18% of participants' ages were in the range from 25 to 29 inclusive, 5% were 30 to 34, and 2% were more than 35 years of age. There were 197 participants with a high school degree (12-years of schooling), 117 participants with a bachelor degree (14-15 years of schooling), and 98 participants had a masters' degree in their respective subject.

Item-Level Content Validity of the TSES

Content validity is the degree to which an item relates to the targeted objective for measuring a specific construct (Haynes, Richard, & Kubany, 1995). Content validity was evaluated by the Index of Item-Object Congruence (IIOC) procedure recommended by Rovinelli and Hambleton (1977). These IIOC values range from -1 to +1. The cutoff scores for IIOC to distinguish "good" items from "not good" items is based on Rovinelli and Hambleton's (1977) recommendations. Content validity is established, when at-least three-quarter of the experts rate an item to be a clear measure of the proposed objective and not a measure of the invalid

objectives. Thus, the index value of .75 would indicate that at least 75% of the experts assign an item a perfect rating.

The Index of Item-Object Congruence values for all 24 items from the Teacher Sense of Efficacy Scale ranged from .68 to 1.00 (see Table 7). The average IIOC value for all 24 items was 0.91 (SD = .11). Item 22, *"How much can you assist families in helping their children do well in school?"* had the lowest IIOC value (IIOC = .68) due to judges' indicating that this was not a clear measure of the *Student Engagement* objective and might be an indirect measure of student engagement. All other items had IIOC values $\geq .72$. According to the histogram (Figure 4), 16 items had IIOC values in between .88 to 1.00, while seven items had values ranging from .72 to .88. The average index values for the construct of *Student Engagement*, *Classroom Management*, and *Instructional Strategies* were .83, .97, and .93, respectively. The item-level indices show that content experts believed 23 out of 24 items of the TSES are associated with the three proposed objectives of the instrument (i.e., *student engagement*, *instructional strategies*, and *classroom management*) with one item scoring in the marginal range due to its match with scale content. No items were identified as potentially problematic due to their interpretation by a Pakistani teacher population.

Descriptive Statistics of the TSES

Descriptive statistics for the three scales of the short and long forms of the TSES are reported in Table 8 for the two sample subgroups. The average efficacies measured by the short form (12-item) TSES for the in-service teachers in the area of *Student Engagement*, *Instructional Strategies*, and *Classroom Management*, were 7.53, 7.48, and 7.67 respectively. The total TSES average (general teaching efficacy) was 7.56. Efficacies measured by the long-form (24-item) for

the in-service teachers in the efficacy of *Student Engagement*, *Instructional Strategies*, and *Classroom Management*, were 7.49, 7.49, and 7.54 respectively.

As far as pre-service teachers' efficacy is concerned, the average efficacies measured by the short form (12-item) TSES for pre-service teachers in the area of *Student Engagement*, *Instructional Strategies*, and *Classroom Management*, were 7.08, 6.87, and 7.07 respectively, with an overall general teaching efficacy average of 7.01 (see Table 8). The average efficacy scores measured by the long form (24-item) for pre-service teachers in the area of *Student Engagement*, *Instructional Strategies*, and *Classroom Management* were 6.81, 6.89, and 6.95 respectively. The overall general teaching efficacy of the pre-service teacher in the long-form was 6.88. These mean values indicate that in general pre-service teachers reported slightly higher efficacy measured by the short-form TSES than those measured by the long-form TSES.

Reliability Analysis

The internal consistency reliability was calculated using Cronbach's (1951) alpha. The reliability coefficients of the overall TSES of all participants for short and long forms were .84 and .92, respectively. These reliability coefficients indicate moderately high levels reliability for the instrument as a whole. Henson (2001b) recommended a reliability coefficient of .80 as an acceptable criterion in general research. The reliability coefficients for the three subscales on the short-form TSES: *Student Engagement*, *Instructional Strategies*, and *Classroom Management* were .70, .73, and .70, respectively. While the reliability coefficients of the three subscales in the long-form TSES: *Student Engagement*, *Instructional Strategies*, and *Classroom Management*, were .83, .84, and .83, respectively. The smaller number of items in the short-form subscales compared to the long-form subscales is the most likely reason for the lower reliability

coefficients. Internal consistency coefficients for the three subscales of the short-form and long-form of the TSES are presented separately for pre-service teachers, in-service teachers, and for all participants in Table 9. Reliability levels for the three subscales are similar for in-service and pre-service teacher with slightly higher reliability for the *Classroom Management* efficacy subscale for in-service teachers.

Scale-Level Correlational Analysis

Scores on the short and long form TSES were very similar for in-service teachers. The correlations between the short-form and the long-form scores of the total TSES score was .96. At subscale level, the correlations between the short and long form for *Student Engagement*, *Instructional Strategies*, and *Classroom Management*, were .92, .93, and .92 respectively. This high correlation between the short and long forms of the TSES also demonstrates the relational similarity between the extra items on the long-form and the shared items on the short and long forms. Tyrer et al. (2010) concluded that when the correlation between the short and long forms is extremely high then both forms can be interchangeable. The only consideration is the lower reliability of the scores on the short form.

In terms of in-service teachers, the inter-scale correlations ranged from .49 to .55 on the short-form and .57 to .62 on the long-form (see Table 10). While for pre-service teachers the inter-scale correlations ranged from .39 to .43 on the short-form and .49 to .56 on the long-form TSES. All three-subscales were strongly correlated with the total TSES score for both in-service and pre-service teacher groups.

Confirmatory Factory Analysis (CFA)

CFA (Confirmatory Factor Analysis) was used under a structure equation modeling framework to examine the theoretical and hypothesized models of the TSES for in-service and pre-service teachers proposed by Tschannen-Moran and Woolfolk Hoy (2001). There were four models tested.

1. Model I (Three-factor model for in-service teachers – long-form TSES)
2. Model II (Three-factor model for in-service teachers - short-form TSES)
3. Model III (One-factor model for pre-service teachers - long-form TSES)
4. Model IV (One-factor model for pre-service teachers - short-form TSES)

CFA (Confirmatory Factor Analysis) was conducted with Maximum Likelihood (ML) estimation in the EQS 6.2 software to test the hypothesized models. Current subsample data for the three subscales are considered normally distributed based on skewness and kurtosis values (within a range of ±1). Lomax and Schumacker (1996) suggest that normal theory should be used when dealing with categorical variables when data is with-in ±1 for skewness and kurtosis.

As latent variables cannot be directly measured, the latent variables of *Student Engagement*, *Instructional Strategies*, and *Classroom Management* do not have a unit of measurement. Therefore, the general practice is to fix the latent variable's measure to the first indicator variable which is specified by the factor loading of the indicated variable being set to one (Cheung & Rensvold, 2002; Greyling, 2006). EQS codes for all four models are provided in appendices K, L, M, and N, respectively.

There are several fit indices to evaluate the CFA model. Since the chi-square test is sample size sensitive, it is less useful to assess CFA model fit by chi-square tests. Several robust fit indices were examined including NFI (Normed Fit Index), NNFI (Non-Normed Fit Index; Tucker & Lewis, 1973), CFI (Comparative Fit Index; Bentler, 1990), SRMR (Standardized Root Mean Square Residual; Hu & Bentler, 1995), and RMSEA (Root Mean Square Error of Approximation; Steiger, 1990) to evaluate the degree of data model fit, as recommended by researchers such as Hu and Bentler (1999), Kline (1998), and Bollen and Long (1993). A summary of the criteria used to evaluate the goodness-of-fit indices of the CFA model is presented in Table 3. Item correlation matrices for the above mentioned models are presented in Table 11 to Table 14.

Model I (Three-factor model for in-service teachers - long-form TSES). The first model included three factors for the Teacher Sense of Efficacy Scale which was hypothesized based on previous studies. Figure 5, shows the path diagram of the three-factor CFA model. The rectangular blocks indicate variables (items) being measured by each factor. The oval shape in the Figure 5 indicates latent variables or factors of the TSES. The latent factors were linked to their corresponding items with single directed arrows which shows that each of the latent factors was unidimensional. All items are designed to measure one of the three specific constructs. These three constructs are assumed to be correlated. These correlations among constructs are indicated by double headed arrows.

The 24 measured variables in the model produced $u = \frac{24(24+1)}{2} = 276$ unique variance and covariance pieces of information. Based on the number of parameters to be estimated and the number of available information in the variance covariance matrix, this model was over-

identified with 249 degree of freedoms in the initial model. As a first step, fit indices (i.e., NFI, CFI, and RMSEA) were assessed to evaluate the proposed model. If the model is accepted then the next step is to examine the estimation of the model parameters. These model fit indices were compared to acceptable model fit criteria provided in Table 3.

The fit indices for a three-factor model were acceptable, $\chi 2\,(249) = 556.63$; $p < .001$; NFI = .89; NNFI = .93; CFI = .93; GFI= .92; AGFI=.91; SRMR=.041; RMSEA = .048 [.042, .053]. RMSEA indicates the residual or unexplained variance. The average absolute standardized residual for the three-factor model was .03, indicating very low residual values between hypothesized model and observed data. This provides evidence that the data model fit well. The inter-factor correlations were moderate to strong between *Instructional Strategies* and *Student Engagement* ($r = .73$, $p < .001$); between *Classroom Management* and *Student Engagement* ($r = .68$, $p < .001$); and between *Instructional Strategies* and *Classroom Management* ($r = .67$, $p < .001$).

The standardized and unstandardized factor loadings including standard error as well as R-squared for each parameter of the final model I, are presented in Table 15. Parameter estimations indicated that all 24 items loaded into three dimensions of the TSES. All parameter estimates were significant at the .01 level. All 24 items fit well and factor loadings were larger than recommended criteria $\geq .30$ by Brown (2006) and $\geq .50$ by Byrne (2006). Though item-22, *"How much can you assist families in helping their children do well in school?"* had the lowest factor loading (.49), overall all items fitted well. In general, loadings between .30 and .59 are considered as moderate while $\geq .60$ are considered a high factor loading (Cokluk et al., 2010). The average of all standardized factor loadings was .63. The standardized loadings ranged from

.49 to .66 for *Student Engagement* factor; from .59 to .72 for *Instructional Strategy* factor; and from .56 to .76 for *Classroom Management* Factor. Figure 6 provides the standardized loadings and error variances of the final CFA model I.

The interpretation of the unstandardized factor loadings are similar to the unstandardized regression coefficient (β) in multiple regression. For example, one unit increase in item 4 *"How much can you do to motivate students who show low interest in school work?"* causes a .98 unit increases in the *Student Engagement* factor after holding all other items constant. Generally, unstandardized loadings are less meaningful as it sometime leads to wrong conclusions. Therefore, the most useful way to interpret the coefficient is by standardizing its unit (i.e., $\mu = 0$ and $\sigma = 1$). The CFA standardized factor loading could also be interpreted in a similar way as standardized regression coefficients (*B*) in multiple regression. Thus, one standard deviation change in item 4 would influence a .61 standard deviation change in *Student Engagement* factor after holding all other items constant. R-squared describes the amount of variance that can be explained by a particular variable. For example, 46% of the proportion of variance in item 17 is explained by the factor *Instructional Strategies*. The association between all the estimations and their corresponding latent variable could be interpreted in a similar way. According to Table 15, it can be concluded that the highest effects of item on each factor were TSES 1 and TSES 6 (r^2 = 43%) on *Student Engagement* factor, TSES 23 (r^2 = 52%) on *Instructional Strategies* factor, and TSES 16 (r^2 = 58%) on *Classroom Management* factor.

Model II (Three-factor model for in-service teachers - short-form TSES). The second model included three factors of the short-form Teacher Sense of Efficacy Scale which were hypothesized based on previous studies. The short-form TSES consists of 12 items. These 12

measured variables in the model produced $u = \frac{12(12+1)}{2} = 78$ unique variance and covariance information with 27 free parameters. Based on the number of parameters to be estimated and number of available information in the variance covariance, this model was over-identified with 51 degree of freedom in the model.

The fit indices for the three-factor model for the short-form (12-item) TSES were acceptable under the criteria given in Table 16. The fit indices for model II were, $\chi 2(51) = 119.95$; $p < .001$; NFI = .94; NNFI = .95; CFI = .96; GFI = .97; AGFI = .95; SRMR = .035; RMSEA = .05 [.038, .061]. These fit indices indicate that the model fit the data well. The average absolute residual was .045 and average absolute standardized residual was .02. Thus, this low residual value between hypothesized model and observed data indicated that that the 3-factor model for the short-form TSES fit the data well for in-service teachers. Since, the data model fits well under the given criteria for an acceptable model; there is no need to make any re-specification in the model. Figure 7 provides the standardized loadings and error variances of the final CFA model II. The inter-factor correlations were moderate to strong between *Instructional Strategies* and *Student Engagement* ($r = .76$, $p < .001$); between *Classroom Management* and *Student Engagement* ($r = .68$, $p < .001$); and between *Instructional Strategies* and *Classroom Management* ($r = .63$, $p < .001$).

Table 16 shows the standardized and unstandardized factor loadings including standard error as well as R-squared for each parameter of the final model II. Parameter estimations indicated that all 12 items loaded into three dimensions of the short-form TSES. All parameter estimates were significant at the .01 level. All 12 items fit very well and factor loadings were more than recommended criteria $\geq .30$ by Brown (2006) and $\geq .50$ by Byrne (2006). Though item

12 (t11), *"How much can you do to foster student creativity?"* had the lowest factor loading (.49), overall all items fit well. The average of all 12 standardized estimations was .65. The standardized loadings were ranged from .49 to .64 for *Student Engagement* factor; from .63 to .68 for *Instructional Strategy* factor; and from .68 to .75 for *Classroom Management* Factor. Parameter estimations of the short-form are consistent with parameter estimations of the long-form. According to the variance explained by each parameter (see Table 16), it can be concluded that 56% of the variance in the item *"How well can you establish a classroom management system with each group of students?"* (# 8 in short and # 16 in the long-from) is explained by the indicated factor (*Classroom Management*). The items most correlated with each factor were TSES 4 (r^2 = 54%) on *Student Engagement* factor, TSES 23 (r^2 = 46%) on *Instructional Strategies* factor, and TSES 16 (r^2 = 56%) on *Classroom Management* factor.

Model III (One-factor model for pre-service teachers - long-form TSES). Model III was based on a one-factor model for pre-service teachers proposed by Tschannen-Moran and Woolfolk Hoy (2001). The fit indices did not indicate acceptable model fit according to the criteria provided in Table 3. The fit indices for model III were, $\chi 2(252) = 1103.25$, $p < .001$; NFI = .69; CFI= .74; GFI = .78; SRMR = .08; RMSEA = .09 (.08, .10). The residual was greater than the .06 or less criteria. Acceptable model fit indicated by CFI index value ≥ .90 (Hu & Bentler, 1999) was also not sufficient.

The factor structure did not fit the data even after allowing free estimation of some of the error covariances through the Lagrange multiplier (LM) test. Since the factor structure was not satisfied according to the hypothesized model, no further CFA investigation is required. Therefore, it is less meaningful to examine individual model parameter's estimations. As a next

step, the researcher proceeded to Exploratory Factor Analysis (EFA) to better explore the construct of the TSES for pre-service teachers. Suhr (2006) also suggested that if the fit indices indicate an unacceptable model fit and the hypothesized factor structure cannot be established, an Exploratory Factor Analysis (EFA) is the next step.

Model IV (One-factor model for pre-service teachers - short-form TSES). Model IV was based on the short-form TSES 1-factor model for pre-service teachers. The fit indices did not indicate acceptable model fit, $\chi 2(54) = 344.15$, $p < .001$; NFI = .73; CFI= .76; GFI = .86; SRMR = .08; RMSEA = .11 (.10, .12). RMSEA was above the acceptable criteria. Thus, the hypothesized model is rejected. The model did not fit well even after allowing some free error covariances. The 1-factor model for the short-form TSES did not fit well as proposed by Tschannen-Moran and Woolfolk Hoy (2001) for pre-service teachers. After reviewing the results of model III (one-factor for long-form TSES), the similar results for the model IV were expected, as both models were tested on pre-service teachers' responses.

Exploratory Factor Analysis

Since both CFA models (III & IV) were rejected, the next step was to proceed with an EFA to better explore conceivable latent traits of the Teacher Sense of Efficacy Scale for pre-service teachers.

EFA for pre-service teachers (long-form TSES). To extract the possible number of factors for the 24-item TSES through exploratory factor analysis for pre-service teachers, the maximum likelihood procedure was used with promax (oblique) factor rotation, allowing for correlated factor loadings given that there were both hypothesized and research-substantiated

relationships among factors. Oblique rotation allows for correlated factors without any restriction and the estimation of factor correlations (Muthen & Muthen, 2004). The EFA was conducted using the Maximum Likelihood (ML) estimation method to extract factors. ML method is an iteration process and provides generally more accurate parameter estimations. The prior communality estimation method was set to Squared Multiple Correlations (SMC) to adjust diagonals of the correlation matrix. Squared multiple correlation method for prior communality estimation is used as communality estimates on the matrix diagonals.

The EFA results produced three-factors for the 24-item TSES for pre-service teachers using the extraction method of Kaiser-Guttman rule (eigenvalues > 1). The scree plot (see Figure 8) also indicated three factors could be the possible solution for pre-service teachers. Total preliminary eigenvalues was 16 with an average of .67. Preliminary eigenvalues for the initial three factors were 11.82, 2.43, and 1.75.

Each factor explained 74%, 15%, and 11% of the variance in the 24-item TSES long-form. A factor loading value of .36 was selected as the statistical criterion for retaining an item in a scale generated by SAS (Statistical Analysis System). The results from exploratory factor analysis are provided in Table 17. The 3-factor solution was moderately effective in accounting for the variability in individual item responses, with a range of 32% to 57% of the item-level variability explained by the common factors. According to table 17, the factor loadings (standardized regression coefficients) ranged from .45 to .82 on *Instructional Strategies* scale (factor 1); .41 to .76 on *Student Engagement* scale (factor 2); and .44 to .67 on *Classroom Management* scale (factor 3). None of the items significantly loaded on more than one factor. However, some loadings were partially shared with other factors. The correlation between factor

1 (IS) and factor 2 (SE) was .50; between factor 1 (IS) and factor 3 (CM) was .53; and between factor 2 (SE) and factor 3 (CM) was .56.

EFA for pre-service teachers (short-form TSES). Another exploratory factor analysis was conducted for the short-form TSES (12-item). Consistent with the long-form's factor structure, the short-form TSES also provided similar results with three-factors retained for pre-service teachers using the Kaiser-Guttman rule (eigenvalues > 1). The scree plot (Figure 9) also indicated three factors can be extracted for pre-service teachers. Preliminary eigenvalues for the initial three factors were 4.07, 1.35, and 1.27. Each factor explained 34%, 11%, and 11% of the variance in the 12-item TSES. A factor loading value of .43 was selected as the statistical criterion for retaining an item in a scale. The factor loadings of each item for the short-form TSES are provided in Table 18. The three-factor solution was moderately to highly effective in accounting for the variability in individual item responses, with a range of 49% to 65% of the item-level variability explained by the common factors. According to Table 18, the factor loadings (standardized regression coefficients) ranged from .62 to .84 on *Instructional Strategies* scale (factor 1); .66 to .78 on *Classroom Management* scale (factor 2); and .66 to .79 on *Student Engagement* scale (factor 3). None of the items significantly loaded on more than one factor or on a different hypothesized factor. The inter-factor correlation between factor 1 (IS) and factor 2 (CM) was .37; between factor 1 (IS) and factor 3 (SE) was .38; and between factor 2 (CM) and factor 3 (SE) was .35.

Multitrait-Multimethod Matrix (MTMM) Results

Convergent validity was assessed by investigating the correlation between primary scales and alternative scales that measure similar constructs. The three alternative measures selected to

provide convergent validity were based on their sets of items measuring constructs similar to the items on the TSES scale given that operational definitions were not provided to define the constructs for each of the six subscales. In an MTMM matrix, convergent validity is established when two similar constructs measured using different methods have a high correlation while divergent validity is established when the correlation between two scales measuring a slightly different, but related construct is observed to be lower than the convergent validity coefficients (Campbell & Fiske, 1959). The three alternative scales were the *Instructional Management Scale* (IMS; Martin & Sass, 2010) to relate with *Student Engagement*, *Personal Teaching Efficacy* (PTE; Gibson & Dembo, 1984) to relate with *Instructional Strategies*, and *Behavior Management Scale* (BMS; Nie, Lau & Liau, 2012) to relate with *Classroom Management* subscale.

The MTMM matrix with raw correlations (without correction for attenuation) is presented in Table 19. The reliability coefficient of each scale is presented on the main diagonal in italicized text within parentheses. Convergent validities (monotrait-heteromethod coefficients) appear as bolded font. Correlations for different traits using the same method (heterotrait-monomethod coefficients) are underlined, whereas the correlations among different scales and different methods (heterotrait-heteromethod coefficients) are presented in plain text.

All internal consistency reliabilities fulfilled the acceptable criteria for reliability coefficients ($\geq .80$). The average reliability coefficient for the three subscales of the TSES (method 1) was .83 while the average reliability coefficient for the three alternative scales (method 2) was .82. The Pearson product-moment correlation coefficients among the three subscales of TSES that represent heterotrait-monomethod values were relatively strong, ranging

from .57 to .62. The inter-scale correlations among IMS, PTE and BMS were also positive but relatively smaller ranging from $r = .36$ to .49. The convergent validity between the *Instructional Management Scale* (IMS) and *Student Engagement* (SE) was .42 ($p < .001$; see Table 19), the convergent validity between *Personal Teaching Efficacy* (PTE) and *Instructional Strategies* (IS) was .43 ($p < .001$), and the *Behavior Management Scale* (BMS) and *Classroom Management* (CM) was .41 ($p < .001$).

The divergent validity coefficients for *Student Engagement* (SE) were .40 and .41 from the heterotrait-heteromethod scales and .57 to .62 from the heterotrait-monomethod scales (other scales within the TSES). The divergent validity coefficients for *Instructional Strategies* (IS) were .37 and .39 from the heterotrait-heteromethod scales and .58 to .62 from the heterotrait-monomethod scales. The divergent validity coefficients for *Classroom Management* (CM) were .32 and .36 from the heterotrait-heteromethod scales and .57 to .58 from the heterotrait-monomethod scales. In order to establish convergent and divergent validity, the monotrait-heteromethod correlations (convergent validity) should be higher than heterotrait-monomethod and heterotrait-heteromethod correlations. The lower heterotrait-heteromethod correlations as compared to the monotrait-heteromethod convergent validity coefficients support the original hypotheses. However, although the convergent validity coefficients were higher than heterotrait-heteromethod correlations for all three subscales of the TSES, the correlations between the heterotrait-monomethods were higher than the convergent validity coefficients. These results do not provide strong support for the convergent validity of the three scales with the use of method being a higher correlation between the three TSES scales than matched traits across different methods. The correlations corrected for measurement error were calculated in order to better evaluate the correlations among the constructs estimated without measurement error impacts (see

Table 20). The three construct validity coefficients ranged from .49 to .52. However, the trend in the comparison of convergent validity coefficients to divergent validity coefficients remained the same due to the similarity in reliability coefficient values. Convergent validity coefficients were higher than divergent validity coefficients measure using different methods, but lower than divergent validity coefficients measure using the same method. Thus, the item format within the TSES scale along with correlated constructs resulted in higher correlations internal to the instrument than correlations across instruments.

Multivariate Analysis of Variance for Three Subscales of the TSES

Results from the content validity section (in which item validity was measured) along with the mathematical evidence provided by internal consistency reliability and construct validity (measured using factor analysis) support the hypothesized definitions provided by the TSES authors. The MTMM analyses do not provide strong support for the construct validity of the definitions for the scales. However, based on the psychometric analyses of the groupings of the items being essentially unidimensional and distinct, the TSES scales will be compared using the demographic groupings selected for a more detailed understanding of subgroup differences related to teacher demographics. Several multivariate analyses of variance (MANOVA) were conducted to examine the group mean differences for pre-service and in-service teachers demographics on the set of three Teacher Sense of Efficacy Scales (i.e., SE, IS, and CM). The MANOVA comparisons used the 24-item long form for all TSES scales. An omnibus alpha level of .05 was selected for the multivariate studies in an effort to place a moderate control of overall type I error.

Prior to analyzing the mean differences using MANOVA, the univariate and multivariate model assumptions of normality and homogeneity of variance were tested. Data were also assessed for univariate and multivariate outliers. The following steps were taking before analyzing the MANOVA to make sure that data satisfy the MANOVA assumptions.

Assessing outliers. Outliers are influential data points that can sharply increase or decrease means score. The sharp increase in mean value affects the deviation scores, covariance, and variability. There are certain methods to detect outliers including descriptive and visual methods. The histogram and the QQ plot were used to view any outlier. The IQR (Interquartile Range) method was also used to assess the influential points. This method is defined as,

$$Lower\ range = \ First\ Quartile - 1.5 \times \ (Interquartile\ range)$$

$$Upper\ range = \ Third\ Quartile + 1.5 \times (Interquartile\ range)$$

Any observation beyond the lower and upper values is considered an outlier. It is possible that an observation has a lower range influential point on one variable and the same observation may have an upper range influential point on another variable. Therefore, if that was the case, the data were further assessed side by side at each variable level. If the same observation has influential points in the same direction on all variables, then that observation was deleted from the analysis. The Mahalanobis distance procedure was used to detect multivariate outliers. A multivariate outlier is a particular case in a set of observations via the combination of two or more than two variable scores. SAS's Interactive Matrix Language (IML) was used to calculate the Mahalanobis distance and plotted in the QQ plot to visually assess the multivariate outliers.

Assessing normality. Univariate and multivariate normality were assessed by descriptive values and viewing the data points on the graph. The descriptive values of the skewness and kurtosis provided information about the shape of the distribution of the variable, while histogram and Q-Q plots were used to visually view the shape of the distributions. SAS provides the univariate normality tests statistics values such as Shapiro-Wilk (Shapiro & Wilk, 1965) statistics and Kolmogorov-Smirnov test (Chakravarti, Laha, & Roy, 1967). The Shapiro-Wilk statistic is the ratio of the unique estimator of the variance to the usual corrected sum of squares estimator of the variance. The Shapiro-Wilk value is positive and always ≤ 1, with a value close to one indicating data fit close to the normal distribution (Refaat, 2007). According to the SAS manual, the Shapiro test is better for sample sizes less than 2000, and the Kolmogorov-Smirnov test should be used for sample sizes more than 2000. Multivariate normality was assessed by Mardia's (1970, 1974) skewness and kurtosis chi-square tests and also by QQ plots. A SAS macro was programmed using Interactive Matrix Language (IML) to calculate Mardia's skewness and kurtosis values for the multivariate normality test.

Homogeneity of variance. Brown and Forsythe's (1974) test was used for the univariate homogeneity of variance model assumption. Brown and Forsythe's (1974) test is the robust version of Levene's (1960) test of homogeneity of variance. This homogeneity of variance test does not assume that the population data is normally distributed. Brown and Forsythe's equal variance test is recommended when the responses are ordinal and the normality assumption is not expected (NCSS, 2015). One of the assumptions of the MANOVA is that the within-group covariance matrices are equal, which is also called homogeneity of covariance (HOC). Box's (1949) test is used to assess whether two or more than two covariance matrices are equal. Box (1949) formulated a test statistic based on the likelihood-ratio, called Box's M statistic. When

Box's M test indicates HOC is violated, then Pillai's trace criterion is a more appropriate choice for test statistics (Tabachnick & Fidell, 2013).

Multivariate Analysis of Pre-service Teachers' Efficacy

Two separate MANOVAs were conducted for the pre-service teacher sample using the gender and academic qualification variables to compare to prior studies from other countries and cultures. An alpha level of .05 was selected to control type I error resulting from two analyses for the pre-service teachers group. There were some potential univariate and multivariate outliers. One from the male group and seven from the female group. The outliers were set aside from the datasets to further investigate the MANOVA assumptions. The univariate normality tests were tenable based on skewness and kurtosis values. Skewness and kurtosis values ranged from -0.15 to -0.41 and -0.36 to -0.94 for male participants while -0.51 to -0.86 and -0.41 to 0.48 for female participants. The data was under ±1 kurtosis and skewness values which suggest that data can be treated as normally distributed data. The multivariate normality test was also acceptable using Mardia's kurtosis value ($\beta = 0.99$, $p = .32$) for the male group and ($\beta = 1.8$, $p = .06$) for the female group. The homogeneity of variance model assumption for *Classroom Management* and *Instructional Strategies* was tenable, while the equal variance assumption for the *Student Engagement* variable did not hold based on the Brown and Forsythe's *F* statistics value ($F = 6.61$, $p = .01$). The homogeneity of covariance assumption was also violated. The larger group (female teachers) had the lowest generalized variance ($|S| = 1.08$) and smaller group (male teachers) had the larger generalized variance ($|S| = 1.45$), resulting in a liberal test statistic.

A one-way MANOVA revealed a significant multivariate main effect at $\alpha = .05$ for gender with the pre-service teachers [Wilks' $\lambda = .96$, $F(3, 412) = 5.73$, $p < .001$], concluding that

there were significant differences between males and females among the set of factors of the Teacher Sense of Efficacy Scale. Only 4% of the total variance in the three subscales of the TSES can be accounted for by gender of the pre-service teachers. When the homogeneity of covariance assumption is violated, Pillai's trace criterion is a more appropriate choice for test statistics (Tabachnick & Fidell, 2013). However, with only two groups, Pillai's trace produces identical results. The post-hoc univariate analyses were conducted using a Bonferroni correction for alpha of $\frac{.05}{3} = .017$. The univariate F-test for post-hoc follow-up analysis at .017 indicated that two efficacy subscales: *Student Engagement,* and *Classroom Management* were significantly different, $F(1, 414) = 13.72$, $p < .001$, partial $\eta^2 = .03$ and $F(1, 414) = 11.41$, $p < .001$, partial $\eta^2 = .03$, respectively. However, there was no significance difference between pre-service male and pre-service female teachers in the efficacy of the *Instructional Strategies, F*(1, 414) = 2.60, $p = .11$, partial $\eta^2 = .01$. Relatively, gender had small relationships with all of the efficacy values. The means and standard deviations for males and females on the three indicators of teaching efficacy are reported in Table 21. Female pre-service teachers reported higher efficacy on all three measures: *Student Engagement* (98.3% *CI* = [0.17, 0.81], *d* = 0.41), *Instructional Strategies* (98.3% *CI* = [-0.10, 0.53], *d* = 0.18) and *Classroom Management* (98.3% CI = [0.12, 0.73], *d* = 0.37).

A MANOVA was also used to test the multivariate effect among three academic level groups (i.e., High school, Bachelors, and Masters) in the set of three subscales. Data were assessed for potential univariate and multivariate outliers. Seven cases were set aside due to substantially departing from univariate and multivariate normality. The univariate and multivariate normality assumptions at subscale levels and subgroup levels were satisfied with these outlier removed. The homogeneity of variance test for the *Student Engagement* variable

satisfied the assumption. However, the homogeneity of variance for the *Instructional Strategies* and the *Classroom Management* were also violated. The largest group had the lowest variance and lowest group had the largest variance, resulting in a liberal *F* statistic. The homogeneity of covariance model assumption was also violated with a similar trend in sample size to covariance matrix relationship. A one-way MANOVA indicated that there was a significant multivariate main effect in the academic qualifications of the prospective teachers [Pillai's trace = .06, F (6, 824) = 4.50, $p < .001$], concluding that there was a significant difference among the three subscales of the TSES based on academic qualification of the pre-service teachers.

The pairwise MANOVAs as a follow-up prior to the univariate analyses were conducted for three groups as suggested by Stevens (2002). Three possible pairwise MANOVAs were tested at alpha .017. There was a significant multivariate main effect in the three subscales of the TSES between pre-service teachers with a high school qualification and pre-service teachers with a bachelor's degree [Wilks' $\lambda = .94$, F (3, 305) = 7.57, $p < .001$]; and between high school qualification and masters qualification (Wilks' $\lambda = .95$, F (3, 296) = 5.13, $p = .002$). There was no significant multivariate effect between pre-service teachers with a bachelor's degree and pre-service teachers with a masters' degree [Wilks' $\lambda = .99$, F (3, 219) = 1.00, $p = .40$].

Post-hoc univariate pairwise comparisons were conducted using Tukey-Kramer's *t* test comparing only the high school qualification group to the bachelor's and masters' groups.

The Bonferroni (Dunnett) correction alpha .017 was used for all pairwise comparison to control type I error. There were significant pairwise group mean differences between pre-service

teacher with high school qualification and pre-service teachers with bachelor qualification in the efficacy of *Student Engagement* (98.3% CI = [0.12, 0.80], *d* = 0.38), in the efficacy of *Instructional Strategies* (98.3% CI = [0.015, 0.65], *d* = 0.30), and in the efficacy of *Classroom Management* (98.3% CI = [0.30, 0.95], *d* = 0.55). There were also significant pairwise group mean differences between pre-service teacher with high school qualification and pre-service teachers with master's qualification in the efficacy of *Student Engagement* (98.3% CI = [0.00, 0.69], *d* =0.29), and in the efficacy of *Classroom Management* (98.3% CI = [0.10, 0.73], *d* = 0.38). It was concluded that pre-service teachers with higher schools' qualification reported significantly higher efficacy in classroom management than pre-service teachers with bachelor qualification and master qualification. None of the other pairwise comparison was significant at .017. The mean and standard deviation of pre-service teachers' efficacy in regard to their academic qualification are reported in Table 22.

Multivariate Analysis of In-service Teachers' Teaching Efficacy

Similar MANOVA procedures were used for in-service teachers to test the significance mean difference in the efficacy of three set of efficacy measures. Prior to analyzing the mean difference between male and female in-service teachers, data were assessed for outlier and assumptions were tested.

The univariate and multivariate normality tests were tenable based on skewness and kurtosis values. The univariate skewness and kurtosis values were under ±1 range which suggests that data can be treated as normally distributed data. Multivariate normality for each group was also acceptable. The homogeneity of variance for *Student Engagement* and *Classroom Management* was tenable while for *Instructional Strategies* variable the equal variance

assumption did not hold. The large group (female teachers) had the smaller variance and small group (male teachers) had the larger variance. The impact of the violation of equal variance for *Instructional Strategies* group leans towards a liberal *F* test statistic. The homogeneity of covariance statistic was also significant. The smaller group had the larger covariance (male = 171, |S| = .35) and the larger group had the smaller covariance (female = 378, |S| = .15), resulting in a liberal *F* test statistic.

A one-way MANOVA (Multivariate Analysis of Variance) revealed a significant multivariate main effect at $\alpha = .05$ in the gender group of in-service teachers [Pillai's trace = .10, F (3, 545) = 20.84, $p < .001$), concluding that there were significant differences among three subscales of the TSES. There was an overall 10% of the total effect accounted for by gender of the in-service teachers in the scores of the three subscales of the TSES. The univariate *F*-test analysis at .017 indicated that there was a significance difference between male and female in-service teachers in the efficacy of *Student Engagement*, $F(1, 547) = 46.49$, $p < .001$, partial $\eta^2 = .08$; *Instructional Strategies*, $F(1, 547) = 42.78$, $p < .001$, partial $\eta^2 = .07$; and *Classroom Management*, $F(1, 547) = 43.71$, $p < .001$, partial $\eta^2 = .07$. Descriptive statistics for male and female in-service teachers in three set of teaching efficacy are reported in Table 23. It is concluded that female in-service teachers have significantly higher sense of efficacy than male in-service teachers in the efficacy of *Student Engagement*, efficacy of *Instructional Strategies*, and efficacy of *Classroom Management* (98.3% *CI* = [0.33, 0.69], $d = 0.63$), efficacy of *Instructional Strategies* (98.3% *CI* = [0.35, 0.75], $d = 0.58$), and efficacy of *Classroom Management* (98.3% *CI* = [0.37, 0.79], $d = 0.61$).

Another one-way MANOVA was used to test the difference in the set of three efficacies for six age groups of in-service teachers. The univariate normality assumption was tenable for each group and each variable. The multivariate normality assumption was also tenable for all six age groups based on the Mardia's skewness and kurtosis values ranging from -0.99 to 1.65 for kurtosis with all p-values > .05. Brown and Forsythe's test satisfied the homogeneity of variance assumption for all three variables (SE, IS, and CM). The homogeneity of covariance assumption using the Box's M test was not tenable. The largest group had the smallest variance (30 to 40 years, $N = 105$, $|S| = 0.10$) and the smallest group had the second largest variance (below 25 years, $N = 66$, $|S| = 0.18$), resulting in a liberal F test statistic. The MANOVA results indicated that there was no significance multivariate main effect for the six age groups of in-service teachers, Wilks' $\lambda = .97$, F (15, 1493.9) = 1.01, $p = .44$. The mean and standard deviations of the three efficacies for in-service teachers are presented in Table 24. It is concluded that there was no significant difference in teaching efficacy among in-service teachers based on age group.

Another one-way MANOVA was used to test the difference in the three set of efficacies in the teaching grade level of in-service teachers. Prior to analyzing the MANOVA, data were assessed for outlier and assumptions were tested. There were seven potential outlier cases causing a violation of normality. These outliers were set aside from the analysis to further investigate the MANOVA assumptions. The univariate and multivariate normality assumptions were tenable for each group. The homogeneity of variance for each variable was also tenable. However, the homogeneity of covariance assumption did not hold using the Box's M test. The largest group had the smallest variance (secondary grades, $N = 175$, $|S| = 0.12$) while the second largest group has the largest variance (elementary grades, $N = 148$, $|S| = 0.26$), making interpretations on the impact to F test statistic difficult. One way MANOVA indicated that there

was a significant multivariate main effect in the teaching grade level of in-service teachers [Pillai's trace = .06, F (9, 1614) = 3.57, $p < .001$] on the three subscales of the TSES.

There were four levels of *teaching grade* variable, therefore six possible pairwise MANOVA were tested at $\frac{.05}{6} = .008$. Three pairwise MANOVA results were significant between in-service teachers teaching at primary level and teachers teaching at elementary level [Pillai's trace = .06, F (3, 262) = 5.97, $p < .001$]; between teachers teaching at primary level and teachers teaching at secondary level [Pillai's trace = .09, F (3, 289) = 9.75, $p < .001$]; and between teachers teaching at primary level and teachers teaching at higher secondary levels [Pillai's trace = .08, F (3, 215) = 6.53, $p < .001$]. Results indicated that there were no significant multivariate difference between teachers teaching at elementary level and teachers teaching at secondary level [Pillai's trace = .01, F (3, 319) = 1.11, $p = .35$]; and teachers teaching at elementary level and teachers teaching at higher secondary level [Pillai's trace = .002, F (3, 248) = 0.15, $p = .93$]; and teachers teaching at secondary level and teachers teaching at higher secondary level [Pillai's trace = .003, F (3, 272) = 1.11, $p = .79$].

Post-hoc univariate pairwise comparisons were conducted using Tukey Kramer's t test. The Bonferroni (Dunnett) correction alpha .017 was used for all pairwise comparison to control type I error.

In-service teachers teaching at primary level reported significantly higher efficacy in *Student Engagement* than teachers teaching at elementary level (98.3% *CI* [.18, .66], $d = 0.52$), in *Instructional Strategies* (98.3% *CI* [.01, .55], $d = 0.31$), and in *Classroom Management* (98.3% *CI* [.03, .62], $d = .33$). Primary teacher efficacy was also significantly higher than teachers teaching at secondary level in *Student Engagement* (98.3% *CI* [.26, .71], $d = 0.61$), in

Instructional Strategies (98.3% *CI* [.21, .71], *d* = 0.53), and in *Classroom Management* (98.3% *CI* [.19, .73], *d* = .48). Primary teacher efficacy was also significantly higher than teachers teaching at higher secondary level (98.3% *CI* [.19, .69], *d* = 0.57), in *Instructional Strategies* (98.3% *CI* [.07, .63], *d* = 0.41), and in *Classroom Management* (98.3% *CI* [.09, .70], *d* = .43). The effect sizes were medium according to Cohen's guidelines for effect size (Cohen, 1988). None of the other pairwise comparisons was significant at .017. The mean and standard deviation of in-service teachers' efficacy in regard to their teaching grade level are reported in Table 25.

In-service teacher's teaching experience ranged from two to 36 years (*M* = 11.80, *SD* = 8. 49). This scale-type response was formed into four teaching experience groups (e.g., 2 – 8 years, 9 – 15 years, 16 – 22 years, and more than 22 years). All the univariate and multivariate normality and homogeneity variance assumptions met the requirements of the MANOVA. One way MANOVA indicated that there was a significant multivariate main effect in the teaching efficacy among teaching experience group [Wilks' $\lambda = .93$, $F(9, 1321.7) = 4.56$, $p < .001$], concluding that there was significant differences among the three subscales of the TSES in the teaching experience groups of in-service teachers.

There were four levels of *teaching experience* variable, therefore six possible pairwise MANOVA were tested at $\frac{.05}{6} = .008$. Three pairwise MANOVA results were significant between teaching experience group 1 (2 - 8 years) and group 3 (16 – 22 years), Wilks' $\lambda = .95$, $F(3, 359) = 6.46$, $p < .001$; group 2 (9 - 15 years) and group 3 (16 – 22 years), Wilks' $\lambda = .90$, $F(3, 205) = 7.15$, $p < .001$; and group 3 (16 – 22 years) and group 4 (22+ years), Wilks' $\lambda = .92$, $F(3,171) = 5.05$, $p = .002$). The other three pairwise MANOVAs were not significant, group 1 (2 - 8 years) and group 2 (9 – 15 years), Wilks' $\lambda = .98$, $F(3, 370) = 2.76$, $p = .04$); group 1 (2 - 8 years) and

group 4 (22+ years), Wilks' $\lambda = .98$, F (3, 336) = 2.50, $p = .06$); and group 2 (9 - 15 years) and group 4 (22+ years), Wilks' $\lambda = .94$, F (3,182) = 3.62, $p = .014$).

When the follow-up tests conducted for significant pairwise MANOVA results, it was found that the efficacy of teachers in group 3 (16 – 22 years) was significantly higher than group 1 (2 - 8 years) in the efficacy of *Classroom Management* (98.3% *CI* = [0.14, 0.70], $d = 0.43$). While the other two efficacy constructs were not significant. Follow-up test for second significant pairwise MANOVA result revealed that teachers in the teaching experience group 3 (16 – 22 years) was significantly higher than group 2 (9 - 15 years) in the efficacy of *Classroom Management* (98.3% *CI* = [0.09, 0.72], $d = 0.43$). No significant differences were found between group 3 and group 2 in the efficacy of *Student Engagement* and *Instructional Strategies*. Teachers at teaching experience group 4 (22+ years) have a higher sense of efficacy in *Student Engagement* and *Instructional Strategies*. Results concluded that higher experience teachers reported higher sense of efficacy than lower experience teachers. The mean and standard deviation of in-service teachers' efficacy in regard to teaching experience groups are reported in Table 26.

Another one-way MANOVA was used to test the difference in the set of efficacies comparing permanent in-service teachers and teachers with temporary contracts. The results revealed that there was no significance multivariate main effect between permanent and contractual in-service teachers, Wilks' $\lambda = .99$, F (3, 545) = 1.66, $p = .17$. According to descriptive statistics presented in Table 26, in-service teachers with permanent job status reported higher efficacy than in-service teachers with temporary job status, however, these differences are not statistically significant.

CHAPTER V

Discussion

Teacher Sense of Efficacy scale is the composition of three subscales: efficacy in student engagement, efficacy in instructional practices, and efficacy in classroom management. There are 8 items and 4 items in each subscale of the long-form (24-item) and short-form (12-item) of the TSES, respectively. The primary purpose of this study was to explore the psychometric properties of the Teachers Sense of Efficacy Scale (TSES) for Pakistani pre-service and in-service teachers. In addition, this study was designed to examine the validity of the Teacher Sense of Efficacy Scale (TSES) and to determine the constructs structure of the TSES in the context of Pakistani pre-service and in-service teachers. Reliability and several types of validation for the TSES were investigated (i.e., content validity and construct validity, including convergent and divergent validity). Finally, this study examined the relationships between Pakistani teachers' characteristics and their sense of efficacy beliefs by comparing the mean differences of the efficacy beliefs based on teachers' gender, age group, teaching level, professional education, and teaching experience.

This research was divided into two main sections. First, to assess psychometric properties of the TSES items and to examine the latent structure of the TSES. Second, to determine the mean differences in efficacy beliefs based on teachers' characteristics. The first section was divided into four phases: item validity, reliability analysis, factor structure, and construct validity, comparing convergent and divergent validity coefficients. A total of 549 in-service teachers' responses and 423 pre-service teachers' responses were used through-out the data analysis process.

Psychometric Properties of the TSES

There are several ways to evaluate the evidence of validity. Construct validity requires sophisticated internal and external evidence. Internal validity refers to the fundamental properties of a measure such as item content and its relationship to a broader theoretical framework (Wasserman & Bracken, 2003) whereas, external validity is described as measures sharing a consistent relationship with their theoretical expectations (Wasserman & Bracken, 2003). External validity can also refer to discriminant validity, criterion related-related validity, and convergent validity (Morris, 2011).

Content validity was evaluated by content experts. Seven experts reviewed each item thoroughly, and their reviews were quantified using the Index of Item-Objective Congruence method (Rovinelli & Hambleton, 1977). The index values for content validity for all 24 items ranged .68 to 1.00 (M = .91, SD = .11). Item 22, *"How much can you assist families in helping their children do well in school?"* had the lowest IIOC value. The cause of the low rating for this item was due to some judges' opinion that this item might not clearly represent any of the three subscales. However, even this value of .68 provides marginal support for the item. The index values for most of the items were perfectly 1.00. Content validity results indicate the subject experts had consensus that all 24-items aligned with their intended objectives, and none of the items were flagged for wording that might be interpreted differently by a Pakistani teacher population. Thus, the item validity feedback demonstrated support of item-level content validity for the TSES.

Reliability analysis was calculated using Cronbach's alpha. It was hypothesized that TSES would produce adequate internal consistency coefficients at the subscale level. Results

indicated that the TSES has good internal consistency reliability. The long-form TSES has reliability coefficients larger than .80 in all three subscales of the TSES (i.e., Student Engagement, Instructional Strategies, and Classroom Management). On the other hand, the short-form TSES has lower reliabilities ranging from .69 to .86. The fact that the short-form TSES has 12 items with four items in each subscale, can cause lower reliability compared to the long-form TSES with eight items in each subscale. Overall, both the short-form and the long-form TSES produced good total test reliabilities of .84 and .92, respectively for both pre-service and in-service teachers.

The construct validity of the TSES was evaluated through an SEM framework. In this study, four models were proposed based on the theory. Model I was a three-factor solution for in-service teachers for the long-form TSES, Model II was a three-factor model for in-service teachers for the short-form TSES, Model III was a one-factor solution for pre-service teachers for the long-form TSES, and Model IV was a one-factor model for pre-service teachers for the short-form TSES. These confirmatory factor analysis (CFA) models were tested through a structure equation modeling (SEM) framework. There are several fit indices to determine the adequacy of the models. The models were evaluated by the multi-index fit indices criteria recommended by Hu and Bentler (1999), including RMSEA (Root Mean Square Error Approximation), SRMR (Standardized Root Mean Square Residual), NFI (Normed Fit Index), NNFI (Non-Normed Fit Index), and CFI (Comparative Fit Index). This statistical method with a multivariate approach is used to test the predetermined construct and adequacy of the theoretical model with the observed data.

The results for the 24-item TSES indicate that a three-factor model is appropriate for in-service teachers as proposed by Tschannen-Moran and Woolfolk Hoy (2001). The data model fit indices indicated that three-factor models is appropriate for in-service teachers. The average absolute standardized residual for the three-factor model for in-service teachers was .03. This indicates very low discrepancy between the hypothesized model and observed data. This further provides evidence of construct validity. Similar results were produced when evaluating model II (three-factor for in-service teachers for the short-form TSES). Fives and Buehl (2010) study suggested that the three-factor solution for teacher efficacy appears to be more appropriate for in-service teachers. When comparing with one-factor models and three-factor models for the TSES, Yousuf Zai and Soomro (2015) also found that the three-factor correlated structure fit the data well for the TSES.

Nie, Lau, and Liau's (2012) findings also suggested three correlated factors fit the data well. Tschannen-Moran and Hoy (2001)'s theoretical framework and other empirical studies conducted in different countries, indicate that the factorial structure for TSES is relatively stable across culturally diverse settings (Klasssen et al., 2009; Yousuf Zai & Munshi, 2016; Yousuf Zai & Soomro, 2015), specifically for in-service teachers.

The CFA results for the 24-item TSES did not support the one-factor model for pre-service teachers as hypothesized by Tschannen-Moran and Woolfolk Hoy (2001). The fit indices (i.e., CFI, NNFI, RMSEA) were not in the range of model acceptance. Thus, the EFA procedure was used to explore the latent structure of the TSES for pre-service teachers. Suhr (2006) suggested that if the CFA model does not fit the data well, the next step is to proceed with exploratory factor analysis to establish the factor structure. EFA results suggested that a three-

factor would be a more favorable way to operationalize Teacher Sense of Efficacy Scale (TSES) for the pre-service teachers from Pakistan, similar to Pakistani in-service teachers. A three-factor model also fit the data well for the 12-item TSES for the pre-service teacher population. These factor analysis results for the pre-service teachers are different than the proposed factor structure. There may be a difference in Pakistani pre-service teachers compared to pre-service teachers in other countries in that many already have experience in the classroom either in public and private school settings. Another reason could be that pre-service teachers in Pakistan are required to make lesson plans and implement them in public schools as part of the partial degree requirement, similar to some training programs but possibly not all.

Pre-service samples from other studies may have had more students who were earlier in their training program with less or no teaching experience. Prior to 2012, a teaching degree was preferred but not essentially required for Pakistani teachers. Therefore, some pre-service teachers in Pakistan may have some experience in the classroom than pre-service teachers from other countries and with an increased understanding of teaching related tasks, they may be better able to distinguish items related to the three specific constructs of the TSES (*Students Engagement*, *Instructional Strategies*, and *Classroom Management*). As a result, it is recommended that the three efficacy scales of the TSES are appropriate for use for both in-service and pre-service teachers in Pakistan. However, when comparing pre-service teachers in Pakistan to pre-service teachers in other countries, it is also appropriate to use the total efficacy scale score of the TSES when a one-model solution appears more appropriate for some of the groups included in the analysis.

After evaluating the latent structure of the set of scales, the MTMM procedure was used to assess the convergent and divergent validity of the TSES subscales with other measures of the same or similar constructs. MTMM correlation with attenuation and without attenuation were analyzed to better evaluate the true correlations among the constructs.

Campbell and Fiske (1959) explained that there should be stronger correlations among scales of the same constructs measured by different methods than the correlations among different constructs measured by different methods. The current findings support this assumption. Moreover, monotrait-heteromethod correlations should be higher than heterotrait-monomethod correlations. The results of the MTMM do not support this second component of the MTMM assumption. All convergent validity correlations were higher than divergent validity coefficients when compared across instruments (hetero-method). However, the convergent validity coefficients were not higher than the divergent validity coefficients that measured relationships within the same instruments (mono-method) for the TSES scale.

These results are consistent with those of Raykov (2011). In the MTMM, the researcher would finally like to see divergent validity coefficients smaller than convergent validity coefficients, without being uncorrelated. The three subscales of the Teacher Sense of Efficacy (TSES) scale have strong inter-subscale correlations. What was not hypothesized was that the scales within the TSES (SE, IS, and CM) would be more correlated with each other than their convergent validity counterparts indicating that the hetero-trait, mono-method component of the analysis that is desired to be less than the mono-trait, hetero-method was not supported. As a result, there is only partial support for the construct validity of the TSES using the MTMM procedure, because the format in which the items are constructed appear to play a factor in the

relationships among the scales. The results for the TSES only partially support the divergent validity comparisons to the convergent validity relationships with stronger relationships among scales hypothesized with a format (mono-method) factor appearing to impact relationships.

The combination of the reliability and factor structure analyses support the mathematical relationship among the item groupings that represent the three scales of the TSES. The item validity component of the study supports the theoretical design of the three scales, with support that the items measure the constructs intended to be measured. However, it is recommended that operational definitions of scales always be included in scale development in order to allow for a more appropriate evaluation of scale content and construct validity. Last, the MTMM component of the analyses provides only partial support that the theoretical development of the three scales of TSES measure similar constructs to what are in the comparison scales used in the MTMM study. With no operational definitions provided for any of these six scales, selection for the MTMM study was difficult for researchers. However, overall there seems to be moderate support for the use of the three scales as correlated, but unique components of teacher efficacy. Further, the use of the three efficacy scales with a Pakistani teacher population (both in-service and pre-service) appears appropriate given the data obtained.

Teacher Efficacy Beliefs and Teachers' Characteristics

In line with second the phase of the analysis, mean comparisons of efficacy beliefs were tested. TSES functions to measure teachers' sense of efficacy in three elements of teachings: *Student Engagement*, *Instructional Strategies*, and *Classroom Management*. Each element of teaching efficacy is considered as an independent, yet correlated factor. Therefore, a MANOVA

procedure was used to determine the mean difference in efficacy beliefs based on teachers' characteristics.

Results revealed that female in-service teachers in Pakistan have a significantly higher sense of efficacy than male in-service teachers in the efficacy of *Student Engagement*, efficacy of *Instructional Strategies*, and efficacy of *Classroom Management*. Similar results were found by Shane (2010) with in-service teachers in the United States and in Pakistan (Ahmad et al., 2015; Butt et al., 2012). Overall 10% of the total effect in teacher efficacy can be accounted by the gender factor for the Pakistani in-service teachers.

Multivariate analysis revealed that there was a significant main effect in teaching efficacy at different teaching grade level. Primary school teachers tend to have higher efficacies than secondary grade levels teachers in *Student Engagement*, *Instructional Strategies*, and *Classroom Management* with the moderate effect size 0.62, 0.53, and .48, respectively. No significant efficacy difference was found between elementary and secondary school teachers and between secondary to higher secondary school teachers. The data indicate that teachers at primary grade levels appear to be more efficacious in their ability to engage students, manage behaviors, and manipulating instruction strategies than teachers teaching at higher grade levels. Similar results were found in previous research (Fives & Buehl, 2010; Wolters & Daugherty, 2007) that efficacy beliefs are related to the grade level taught by teachers in the United States and in Pakistan (Haq & Akhtar, 2013). They also found that teachers of lower grade levels tend to report higher efficacy levels than teachers of upper higher levels.

Current results indicated that teacher sense of efficacy is significantly related to teacher's teaching experience. Teachers with more years of teaching experience have higher sense of efficacy than teachers with fewer years of teaching experience. Results indicated that

experienced teachers reported a higher sense of efficacy in classroom management than the lower experience teachers. These findings are consistent with Fives and Buehl (2010) findings regarding teaching efficacy and teaching experience of in-service teachers in the United States and in Pakistan (Haq & Akhtar, 2013). Soodak and Podell's (1997) findings also suggested that teachers with more years of teaching experience have higher personal teaching efficacy than teachers with fewer year of teaching experience.

When comparing efficacy beliefs, between pre-service male and pre-service female teachers, there was a significant multivariate effect for the three subscales of the TSES. Results indicated that female pre-service teachers tend to have a significantly higher sense of efficacy than male teachers in the efficacy of *Student Engagement* and in the efficacy of *Classroom Management*. Although there was no significant mean difference between male and female pre-service teachers in the efficacy of *Instructional Strategies*, results indicated that female pre-service teachers reported a higher sense of efficacy in instructional strategies than male pre-service teachers. This result was similar to what was found for in-service teachers in this study.

Limitations of the Study

There are significant benefits of this research as such a large-scale study has been not been conducted before with teachers in Pakistan's context. This is the first validation study of a teacher efficacy measure that contains samples of pre-service and in-service teachers from all major cities of Pakistan in order to provide a relatively large, diverse study for comparison. However, there are limitations of this research to be acknowledged. First, this study is based on a convenience sampling technique. For those groups that were identified for surveying, the response rate was extremely high. However, not all pre-service teachers in Pakistan had the

chance of being sampled. There was also a very small subgroup of the in-service teacher sample that was obtained via online surveys in which the responses were voluntary and a sampling response rate is unable to be calculated. Due to this and other factors, there are certain internal and external validity threats that should be considered while interpreting results. For example, participants with lower efficacy beliefs might not be willing to participate in the study. Second, the instruments used in this study were in English and teachers in some areas in Pakistan are not quite experts in the English language. Therefore, the researcher had to explain each item in a way that participants could understand. The instruments used in this study are in a self-report format, and participants were expected to be honest in their teaching efficacy beliefs. Moreover, participants completing the survey may not be able to describe themselves precisely (Ben-Porath, 2003) which could allow them to be vulnerable to self-belief biases. Alternatively, one might argue that the assessment of teacher efficacy beliefs by self-report should be considered appropriate since there is no expert observer who could have the same specific and accurate information about the individual's efficacy belief as the individual himself (Chan, 2003). Last, one of the measurement limitations is that no formal operational definitions were provided by the construct developer of the TSES nor the comparison scales. This made it very hard to find suitable alternative measures for an ideal MTMM procedure.

Directions for Future Research

There are a few thoughts for future research related to the current study. First, though the TSES is the combination of three useful constructs relating to teaching, the teaching efficacy instrument can be more comprehensive by adding additional factors related to teaching such as efficacy in assessment techniques and efficacy in using teaching aids. Second, although the

TSES was developed to measure constructs for effective teachers (Tschannen-Moran & Woolfolk Hoy, 2001), these constructs can be more subject specific. Bandura (1981) said, efficacy belief is a concept for a specific domain. For example, a teacher may feel a high sense of efficacy when teaching physics, but may not have high efficacy when teaching general mathematics. Future research can be conducted to use the TSES for teachers teaching specific subjects. Third, teacher training institutes in Pakistan recently launched a four-year B. Ed. program that is mandatory for every prospective teacher for secondary schools. Previously, the B. Ed qualification was not mandatory to hire teachers. A candidate with any bachelor degree (i.e., Science, Arts, Math, and English) was able to get a teaching position in public schools while having a professional teacher training degree (i.e., B.Ed, M.Ed) was considered an additional qualification. Since 2012, various teacher training institutes in Pakistan initiated a two-year associate degree in education (ADE) and a four-year bachelor degree in education (B. Ed Hons.) to prepare specialized and qualified teachers. This degree program provides intensive teacher training, pedagogical skills, contents knowledge and opportunity for the internship in public school over a four-year time span. Therefore, measuring the teachers' efficacy over a time period would provide useful information regarding the development of sense of efficacy of prospective teachers. Such longitudinal studies will provide information on different sources that can develop efficacy beliefs. It will also be helpful for institute administrators and policy makers to determine factors that could improve prospective teachers' sense of efficacy. Fourth, teachers' professional development positively influences teachers' efficacy (Tschannen-Moran & McMaster, 2009). Specific professional development programs can be organized to improve in-service teachers' sense of efficacy. As a first step, it is recommended to develop a short teacher training program for in-service teachers to improve teachers' skill sets that can impact sense of

efficacy in three areas of teaching: classroom management, instructional strategies, and student engagement to measure the teachers' self-efficacy before and after training intervention. However, it is also important to advise researchers to be cautions when selecting the TSES for assessing differences in efficacy due to the high score values for both pre-service and in-service teachers obtained in this study. With the average greater than 7.0 on a 9.0 scale (SD ≅ 1), there will be very little room for significant increases in scores, with the potential for a ceiling effect. It is something the researcher should be aware of. Finally, additional validation studies are encouraged that collect other samples of pre-service teachers to conduct CFA analyses using a three-factor model to compare to the current results.

Recommendations and Implications for Practice

The findings from the current study may have valuable practical implications. For example, researchers can use current findings as evidence of validity and reliability of the TSES to measure teachers' self-efficacy beliefs for in-service and pre-service teachers in Pakistan. Given present findings regarding teachers' self-efficacy, it seems reasonable to further explore the relationship between teachers' self-efficacy and other school outcomes such as job satisfaction, teacher burn-out, and student achievement. According to Bandura (1981), efficacy belief can be a construct specific belief. For example, a math teacher may have a different level of efficacy belief when he or she is assigned to teach physics. Teachers may feel more efficacious to teach a specific subject. Therefore, there would be a positive impact on student learning if a teacher is allowed to teach the subject(s) he or she is most comfortable with. Teacher education programs can be designed in a way to better develop teachers' efficacy beliefs both generally and in specific subjects. Bandura (1997) defined four sources that contribute in

developing self-efficacy beliefs: *mastery experiences, learning through vicarious experiences, physiological and emotional states,* and *social persuasion.* According to Bandura (1997), verbal persuasion can help to boost self-efficacy beliefs. School principals and administrators can help to increase teachers' efficacy beliefs by communicating in an encouraging way when dealing with conflicts. For researchers, it is recommended to use the long-form TSES for better reliability. The full version of the TSES provides more information about teachers' efficacy in the areas measured. The data also support the use of the TSES in studying and comparing Pakistani in-service teachers to teachers in other countries. Pre-service teacher comparisons appear appropriate to comparisons with in-service Pakistani teachers given the similarity in model fit using the three-factors defined by the subscales. This may be different from in-service and pre-service teachers in other countries.

References

Adu, E., Tadu, R., & Eze, I. (2012). Teachers' self-efficacy as correlates of secondary school students' academic achievement in southwestern Nigeria. *Discovery, 2*(4), 8-16. Retrieved July 11, 2015, from http://www.discovery.org.in/discovery/current_issue/v2/n4/A3.pdf

AEPAM (2011). An analysis of educational indicators of Pakistan – 2011 by NEMIS-AEPAM, National Education Management Information Academy of Educational Planning and Management, Ministry of Professional and Technical Training, Government of Pakistan, (AEPAM Publication No 244).

Ahmad, R., Khan, S., & Rehman, S. (2015). Comparative study to investigate the sense of teacher efficacy between male and female teachers. *Asian Journal of Management Sciences & Education, 4*(2), 29-35.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 3*, 317-332.

Akram, B., & Ghazanfar, L. (2014). Self-efficacy and academic performance of the student of Gujrat University, Pakistan. *Academic Research International, 5*(1), 283-290. Retrieved July 11, 2015, from http://www.savap.org.pk/journals/ARInt./Vol.5(1)/2014(5.1-30).pdf

Al-Alwan, A., & Mahasneh, A. (2014). Teachers' self-efficacy as determinant of students' attitudes toward school: A study at the school level. *Review of European Studies, 6*(1), 171-179. doi:10.5539/res.v6n1p171

Al-Khalaileh, H. & Abu-Tineh, A. (2011). Self-efficacy of teachers in Zarqa governorate and its relation to their principals' leadership practices. *Dirasat Journal, 38*(2), 222-237.

Alinder, R. M. (1994). The relationship between efficacy and the instructional practices of special education teachers and consultants. *Teacher Education and Special Education, 17*, 86-96.

Alrefaei, N. (2015). *Teachers' sense of efficacy: Examining the relationship of teacher efficacy and student achievement* (Order No. 3704548). Available from Dissertations & Theses @ University of Arkansas Fayetteville; ProQuest Dissertations & Theses Global. (1690497495). Retrieved from http://0-search.proquest.com.library.uark.edu/docview/1690497495?accountid=8361

Anderson, R., Greene, M., & Loewen, P. (1988). Relationships among teachers' and students' thinking skills, sense of efficacy, and student achievement. *Alberta Journal of Educational Research, 34*(2), 148.

Armor, D., Conroy-Oseguera, P., Cox M., King, N., McDonnell, L., Pascal, A. Pauly, E., & Zellman, G. (1976). Analysis of the school preferred reading programs in selected Los

Angeles minority schools. (REPORT NO. R-2007-LAUSD). Santa Monica, CA: Rand Corporation. (ERIC Document Reproduction Service No. 130 243).

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., Buhr, D., & Crocker, L. (1984). Teachers' sense of efficacy: A self- or norm-referenced construct? *Florida Journal of Educational Research, 26* (1). 29-41.

Ashton, P. T & Webb, R. G. (1986). Teachers' sense of efficacy, classroom behavior, and student achievement. In P. T. Ashton & R. B. Webb (Eds.), *Teachers' sense of efficacy and student achievement* (pp. 125-144). New York: Longman.

Avanzi, L., Miglioretti, M., Velasco, V., Balducci, C., Vecchio, L., Fraccaroli, F., & Skaalvik, E. M. (2013). Cross-validation of the Norwegian Teacher's Self-Efficacy Scale (NTSES). *Teaching and Teacher Education*, 31, 169-78. doi:10.1016/j.tate.2013.01.002

Baldwin, B. (1989). A primer in the use and interpretation of structural equation models. *Measurement and Evaluation in Counseling and Development, 22*(2), 100-112. Retrieved from http://0-search.proquest.com.library.uark.edu/docview/63058946?accountid=8361

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

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). Cambridge: Cambridge University Press.

Bandura, A. (1995). Exercise of personal and collective efficacy in changing societies. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 1-45). New York: Cambridge University Press. Retrieved from https://books.google.com/books?id=ZL7qN4jullUC&dq=editions%3AitTbpOYuAYgC&source=gbs_book_other_versions

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

Bandura, A. (1999). Social cognitive theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 154-196). New York: The Guilford Press.

Bandura, A. (2006). Adolescent development from an agentic perspective. In *Self-efficacy beliefs of adolescents* (pp. 1–43). Greenwich, CT: Information Age.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Greenwich, CT: Information Age.

Barker, K., Yeung, A., Dobia, B., & Mooney, M. (2009). *Positive behaviour for learning: Differentiating teachers' self-efficacy*. Paper presented at the Australian Association for Research in Education Annual Conference, Canberra, Australia.

Ben-Porath, Y. S. (2003). Assessing personality and psychopathology with projective methods. In I. B. Weiner (Ed.), *Handbook of psychology* (pp. 531-552). Hoboken, NJ: John Wiley & Sons.

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

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606. doi:10.1037/0033-2909.88.3.588

Bentler, P. M., & Chou, C. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*(1), 78-117.

Billheimer, B. C. (2006). *Perceived teacher self-efficacy in early childhood settings: Differences between early childhood and elementary education candidates* (Order No. 1458305). Available from ProQuest Dissertations & Theses Global. (305325520). Retrieved from http://0-search.proquest.com.library.uark.edu/docview/305325520?accountid=8361 and http://dc.etsu.edu/etd/2200

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models*. Newbury Park, CA: Sage Publications.

Boomsma, A., & Hoogland, J.J. (2001). The robustness of LISREL modeling revisited. In: Cudeck, R., du Toit, S., & Sörbom, D. (Eds.). *Structural Equation Modeling: Present and Future: A Festschrift in Honor of Karl Jöreskog* (pp. 139–168). Chicago: Scientific Software International.

Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika, 36*(3-4), 317–346.

Brophy, J. E., & Good, T. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–375). New York: Macmillan.

Brown, T. A. (2006), *Confirmatory Factor Analysis for Applied Research*. New York: Guilford Press.

Brown, J. D., & Hudson, T. (2002). *Criterion-referenced language testing*. Cambridge, UK: Cambridge University Press.

Brown, M. B., & Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association, 69*(346), 364–367. http://doi.org/10.2307/2285659

Butt, M., Khan, H., & Jehan, S. (2012). Impact of English teachers' self efficacy beliefs on students' performance. *World Applied Sciences Journal, 20*(7), 1031-1035. doi:10.5829/idosi.wasj.2012.20.07.2684

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. Thousand Oaks: Sage Publications.

Byrne. B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates;

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. doi: 10.1037/h0046016.

Cantrell, P., Young, S., & Moore, A. (2003). Factors affecting science teaching efficacy of pre service teachers. *Journal of Science Teacher Education, 14*, 177-192.

Çapa, Y., Çakıroğlu, J., & Sarıkaya, H. (2005). The development and validation of a Turkish version of Teachers' Sense of Efficacy Scale. *Eğitim ve Bilim (Education and Science), 30*(137), 74-81.

Caprara, G. V., Barbaranelli, C., Steca, P., & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of School Psychology, 44*(6), 473-490. doi:10.1016/j.jsp.2006.09.001

Chakravarti, Laha, and Roy, (1967). *Handbook of Methods of Applied Statistics, Volume I,* John Wiley and Sons, pp. 392-394.

Chan, D. W. (2003). Multiple intelligences and perceived self efficacy among Chinese secondary school teachers in Hong Kong. *Educational Psychology, 23*(5), 521-533

Charalambous, C. Y., Philippou, G. N., & Kyriakides, L. (2008). Tracing the development of preservice teachers' efficacy beliefs in teaching mathematics during fieldwork. *Educational Studies in Mathematics, 67*(2), 125-142. doi:10.1007/s10649-007-9084-2

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cokluk, O, Sekercioglu, G. & Buyukozturk, S. (2010). *Multivariate statistics for the social sciences. SPSS and LISREL applications.* Ankara: Pegem Publication.

Coladarci, T. (1992). Teachers' sense of efficacy and commitment to teaching. *Journal of Experimental Education, 60*(4), 323-337. Retrieved from http://0-search.proquest.com.library.uark.edu/docview/62850118?accountid=8361

Coleman, J., Campbell, E., Hobson, D., McPartlend, J., Mood, A., Weinfield, F., & York, R. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Office of Health, Education, and Welfare.

Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, Fort Worth: Harcourt Brace Jovanovich College Publishers.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. doi:10.1007/bf02310555

Dawis, R. (1987). Scale construction. *Journal of Counseling Psychology, 34*(4), 481-489.

De Stercke, J., Temperman, G., De Lièvre, B., & Lacocque, J. (2014). Echelle de sentiment d'efficacité personnelle des enseignants : Traduction francophone de la Teachers' Sense of Efficacy Scale. Service de Pédagogie Générale et des Médias Educatifs, Université de Mons. Retrieved on December 25, 2015 from https://hal.archives-ouvertes.fr/hal-01145411/file/De_Sterckeetal-EchelleSEPEnseignants(tradTSES).pdf

DeVellis, R. F. (2003). *Scale development: Theory and applications.* Newbury Park: Sage.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method*. New Jersey: Wiles & Sons.

Duffin, L. C., French, B. F., & Patrick, H. (2012). The Teachers' Sense of Efficacy Scale: Confirming the factor structure with beginning pre-service teachers. *Teaching and Teacher Education, 28*(6), 827-834. doi:10.1016/j.tate.2012.03.004

Ebstrup, J.F., Eplov, L.F., Pisinger, C., & Jørgensen, T. (2011). Association between the Five Factor personality traits and perceived stress: Is the effect mediated by general self-efficacy? *Anxiety, Stress, and Coping, 24*(4), 407-419. doi: 10.1080/10615806.2010.540012.24:407–419.

Edwards, J. L. Green, K. E. & Lyons, C.A. (1996, April). *Teacher efficacy and school and teacher characteristics*. Paper presented at the annual meeting of the American Educational Research Association, New York.

Ekstrom, R. (1975). *The relationship of teacher aptitudes to teacher behavior: Beginning teacher evaluation study: Phase II. Final report: Volume VI.* Princeton, NJ: Educational Testing Service.

Ereño, J., & Nunez, K. (2014). What makes a teacher a better teacher? A hierarchical regression analysis of self efficacy and commitment as predictors of faculty members' job performance. *International Journal of Asian Social Science, 4*(4), 492-498. Retrieved June 14, 2015, from http://www.aessweb.com/pdf-files/ijass-2014-4(4)-492-498.pdf

Emmer, E., (1990). A scale for measuring teacher efficacy in classroom management and discipline. *Paper presented at the annual meeting of the American Educational Research Association,* Boston, MA (Revised, 1990).

Emmer, E., & Hickman J. (1990). Teacher decision making as a function of efficacy, attribution, and reasoned action. *Paper presented at the annual meeting of the American Educational Research Association,* Boston, MA.

Fives, H., & Buehl, M. M. (2010). Examining the factor structure of the Teachers' Sense of Efficacy Scale. *The Journal of Experimental Education, 78*(1), 118-134. doi:10.1080/00220970903224461.

French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Manual for kit of reference tests for cognitive ability*. Princeton, NJ: Educational Testing Service.

Gavora, P. (2010). Slovak pre-service teacher self-efficacy: Theoretical and research considerations. *The New Educational Review, 21*(2), 17-30.

Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 40-65). Newbury Park, CA: Sage.

Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76*(4), 569-582. doi:10.1037/0022-0663.76.4.569

Goddard, R. D., Hoy, W. K., & Woolfolk Hoy, A. (2004). Collective efficacy beliefs: Theoretical developments, empirical evidence, and future directions. *Educational Researcher, 33*(3), 3-13.

Greyling. T. (2006). *Validation of a new measuring instrument of quality of life: A confirmatory factor analysis approach*. Retrieve from http://www.researchgate.net/file.PostFileLoader.html?id=55ad0a7d6225ffbb608b45d5&assetKey=AS%3A273816754819073%401442294410656

Guerreiro-Casanova, D. C. & Azzi, R.G. (2013). *Efficacy beliefs and teachers' job satisfaction: A Brazilian study*. In: 35th International School Psychology Conference. Universidade do Porto, Porto, Portugal, 2013b.

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

Guskey, T. R. (1982). Differences in teachers' perceptions of personal control of positive versus negative student learning outcomes. *Contemporary Educational Psychology, 7*(1), 70-80.

Guskey, T.R. (1988). Teacher efficacy, self-concept, and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education, 4*(1), 63-69. doi:10.1016/0742-051X(88)90025-X

Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness & School Improvement, 9*(2), 157-191.

Haq, M., & Akhtar, M. (2013). Do school level and years of teaching experience really matter? An investigation of Pakistani teachers' self-efficacy beliefs. *Middle-East Journal of Scientific Research, 16*(5), 644-656. doi:10.5829/idosi.mejsr.2013.16.05.1290

Haynes, S., Richard, D., & Kubany, E. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7(3),* 238-247.

Henson, R. K. (2001a). The effects of participation in teacher research on teacher efficacy. *Teaching and Teacher Education, 17*(7), 819-836.

Henson, R. K. (2001b). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development, 34*(3), 177-189.

Hollon, R. E., Anderson, C. W., & Roth, K. L. (1991). Science teachers' conceptions of teaching and learning. In J. Brophy (Ed.), *Advances in research on teaching* (pp. 145-185). Greenwich, CT: JAI Press.

Holzberger, D., Philipp, A., & Kunter, M. (2013). How teachers' self-efficacy is related to instructional quality: A longitudinal analysis. *Journal of Educational Psychology, 105*(3), 774-786. doi:10.1037/a0032198

Hoy, W. K., & Woolfolk, A. E. (1993). Teachers' sense of efficacy and the organizational health of schools. *The Elementary School Journal, 93*(4), 355-372. doi:10.1086/461729

Hoyle, R. H. (1995). *Structural equation modeling: Concepts, issues, and applications.* Thousand Oaks: Sage Publications.

Hu, L.-T., & Bentler, P. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural Equation Modeling. Concepts, Issues, and Applications* (pp.76-99). London: Sage.

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

Hughes, A. (2003). *Testing for language teachers*. Cambridge University Press. Retrieved from Cambridge Language Teaching Library: http://dx.doi.org/10.1017/CBO9780511732980

Hunt-Ruiz, H. (2011). *Preservice teachers: Investigations in early fieldwork and mathematics efficacy beliefs*. Available from ERIC. (1037908076; ED534070). Retrieved from http://0-search.proquest.com.library.uark.edu/docview/1037908076?accountid=8361

Janda, L. H. (1998). *Psychological testing: Theory and applications*. Boston: Allyn & Bacon.

Jensen, A.R. (1998). *The g Factor: The Science of Mental Ability*. Praeger, Connecticut, USA. ISBN 0-275-96103-6

Jöreskog, K. G., & Sörbom, D. (1981). *LISREL V: Analysis of linear structural relationships by the method of maximum likelihood*. Chicago: National Educational Resources.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.

Khan, A. (2011). Teacher efficacy - A tool to enhance academic achievement of secondary schools. *Language in India, 11*(6), 235-247.

Khan, S. A. (2012). The relationship between teachers' self-efficacy and students' academic achievement at secondary level. *Language in India, 12*(10), 436-449.

Khezerlou, E. (2013). Teacher self-efficacy as a predictor of job burnout among Iranian and Turkish EFL teachers. *Procedia- Social and Behavioral Sciences,* 70, 1186-1194. http://dx.doi.org/10.1016/j.sbspro.2013.01.175

Klassen, R. M., & Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of Educational Psychology, 102*(3), 741.

Klassen, R. M., Bong, M., Usher, E. L., Chong, W. H., Haun, V. S., Wong, I. F., & Georgiou, T. (2009). Exploring the validity of a teachers' self-efficacy scale in five countries. *Contemporary Educational Psychology, 34*(1), 67-76. doi:10.1016/j.cedpsych.2008.08.001

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: The Guilford Press.

Kunnan, A. J. (1998). An introduction to structural equation modeling for language assessment research. *Language Testing, 15*(3), 295–332.

Levene, H. (1960). *In Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling,* I. Olkin et al. eds., Stanford University Press, pp. 278-292.

Levin, J. D., Culkin, J., & Perrotto, R. S. (2001). *Introduction to chemical dependency counseling*, North Bergen, NJ: Book-mart Press, Inc. Retrieved from http://0-search.proquest.com.library.uark.edu/docview/200012329?accountid=8361

Lewis, T. & Sugai, G. (1999). Effective behaviour support: A systems approach to proactive school wide management, *Focus on Exceptional Children, 31*(6), 1-24.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*(9), 705-717.

Lomax, R. G. (1989). Covariance structure analysis: Extensions and developments. In B. Thompson (Ed.), *Advances in social Science methodology (*pp. 171-204). Greenwich, CT: JAI.

Luszczynska, A., Gutiérrez-Doña, B., & Schwarzer, R. (2005). General self-efficacy in various domains of human functioning: Evidence from five countries. *International Journal of Psychology, 40*(2), 80-89.

Maas, C. M., Lensvelt-Mulders, G. M., & Hox, J. J. (2009). A multilevel multitrait-multimethod analysis. *Methodology 5*(3), 72-77. doi:10.1027/1614-2241.5.3.72

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*, 519–530.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya B, 36*, 115–128.

Martin, N. K., & Sass, D. A. (2010). Construct validation of the behavior and instructional management scale. *Teaching and Teacher Education: An International Journal of Research and Studies, 26*(5), 1124-1135.

Midgley, C., Feldaufer, H., & Eccles J. (1989). Change in teacher efficacy and student self- and task-related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology, 81*(2), 247-258

Millsap, R. E. (2007). Structural equation modeling made difficult. *Personality and Individual Differences, 42*(5), 875-881. doi:10.1016/j.paid.2006.09.021

Minke, K. M., Bear, G. G., Deemer S. A., & Griffen, S. M. (1996). Teacher experiences with inclusive classrooms: Implications for special education reform. *The Journal of Special Education, 30*(2), 152-186.

Mojavezi, A., & Tamiz, M. (2012). The impact of teacher self-efficacy on the students' motivation and achievement. *Theory and Practice in Language Studies, 2*(3), 483-491. doi:10.4304/tpls.2.3.483-491

Moore, W., & Esselman, M. (1992). *Teacher efficacy, power, school climate and achievement: A desegregating district's experience*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Morris, D. B. (2011). Sources of teaching self-efficacy: A scale validation. *Dissertation Abstracts International Section A, 71*, 3898.

Morris-Rothschild, B. K., & Brassard, M. R. (2006). Teachers' conflict management styles: The role of attachment styles and classroom management efficacy. *Journal of School Psychology, 44*(2), 105-121. doi:10.1016/j.jsp.2006.01.004

Mueller, R. O. (1997). Structural equation modeling: Back to basics. Teacher's corner. *Structural Equation Modeling: A Multidisciplinary Journal, 4*(4), 353-69.

Mundfrom, D. J., Shaw, D. G., & Tian Lu, K. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing, 5*(2), 159-168. doi:10.1207/s15327574ijt0502_4

Muthen, L. K. & Muthen, B. O. (2004). *Mplus user's guide* (3rd ed). Los Angeles: StatModel.

NCSS. (2015). Brown-Forsythe Test of Variances (Simulation). Retrieved December 25, 2015, from http://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/PASS/Brown-Forsythe_Test_of_Variances-Simulation.pdf

Nie, Y., Lau, S., & Liau, A. (2012). The teacher efficacy scale: A reliability and validity study. *The Asia-Pacific Education Researcher, 21*(2), 414-421.

O'Neill, S., & Stephenson, J. (2012). Exploring Australian pre-service teachers' sense of efficacy, its sources, and some possible influences. *Teaching and Teacher Education: An International Journal of Research and Studies*, *28*(4), 535-545.

Penrose, A., Perry, C., & Ball, I. (2007). Emotional intelligence and teacher self efficacy: The contribution of teacher status and length of experience. *Issues in Educational Research, 17(*1), 107-126. Retrieved from http://0-search.proquest.com.library.uark.edu/docview/61949351?accountid=8361

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

Quinn, M. (2007). *Assessment and correlates of teaching self-efficacy in a preschool and kindergarten teacher population* (Order No. 3291235). Available from ProQuest Dissertations & Theses Global. (304842558). Retrieved from http://0-search.proquest.com.library.uark.edu/docview/304842558?accountid=8361

Ravert, R. D., Gomez-Scott, J., & Donnellan, M. B. (2015). Equivalency of paper versus tablet computer survey data. *Educational Researcher*, *44*(5), 308. Retrieved from http://0-search.proquest.com.library.uark.edu/docview/1696028462?accountid=8361

Raykov, T. (2011). Evaluation of convergent and discriminant validity with multitrait-multimethod correlations. *The British Journal of Mathematical and Statistical Psychology, 64*(1), 38-52. doi:10.1348/000711009X478616

Refaat, M. (2007). *Data preparation for data mining using SAS*. San Francisco: Morgan Kaufmann.

Rizvi, M., & Elliot, B. (2005). Teachers' perceptions of their professionalism in government primary schools in Karachi, Pakistan. *Asia Pacific Journal of Teacher Education, 33*(1), 35-52. Retrieve from http://ecommons.aku.edu/pakistan_ied_pdck/79

Rose, J. S., & Medway, F. J. (1981). Measurement of teachers' beliefs in their control over student outcome. *The Journal of Educational Research, 74*(3), 185-90.

Ross, J. A. (1992). Teacher efficacy and the effect of coaching on student achievement. *Canadian Journal of Education, 17*(1), 51–65.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs, 80*(1), 1-28.

Rovinelli, R. J., & Hambleton, R. K. (1977). On the use of content specialists in the assessment of criterion-referenced test item validity. *Dutch Journal of Education Research, 2*(2), 49-60.

Ruan, J., Nie, Y., Hong, J., Monobe, G., Zheng, G., Kambara, H., & You, S. (2015). Cross-cultural validation of teachers' sense of efficacy scale in three Asian countries: Test of measurement invariance. *Journal of Psychoeducational Assessment, 33*(8), 769-779. doi:10.1177/073428291557402

Saklofske, D.H., Michayluk, J.O., & Randhawa, B.S. (1988). Teachers' efficacy and teaching behaviors. *Psychological Reports, 63*(2), 407-414. doi:10.2466/pr0.1988.63.2.407

Schaffer, E. C., Muijs, R. D., Kitson, C. & Reynolds, D. (1998). *Mathematics enhancement classroom observation record*. Newcastle upon Tyne: Educational Effectiveness and Improvement Centre.

Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.). New York, NY, US: Routledge/Taylor & Francis Group.

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston, *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35- 37). Windsor, England: NFER-NELSON.

Schwarzer, R., & Jerusalem, M. (2014, October 13). The General Self Efficacy. Retrieved July 11, 2015, from http://userpage.fu-berlin.de/~health/selfscal.htm

Shane, J. A. (2010). *The efficacy of effort differences in teachers' sense of efficacy based on type of teacher training and number of years of experience* (Unpublished doctoral dissertation # CFE0003268). University of Central Florida. Access link http://purl.fcla.edu/fcla/etd/CFE0003268

Shapiro, S. S. & Wilk, M. B. (1965). Testing the normality of several samples. (Unpublished manuscript)

Shaukat, S., & Iqbal, H. M. (2012). Teacher self-efficacy as a function of student engagement, instructional strategies and classroom management. *Pakistan Journal of Social and Clinical Psychology, 9*(3), 82-86.

Skaalvik, E. M., & Skaalvik, S. (2010). Teacher self-efficacy and teacher burnout: A study of relations. *Teaching and Teacher Education, 26*(4), 1059-1069. doi:10.1016/j.tate.2009.11.001

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85*(4), 571-581. Retrieved from https://www.pdx.edu/sites/www.pdx.edu.psy/files/Assessment-11-Motivation-in-the-classroom--reciprocal-effects-of-teacher-behavior--Skinner-Belmont--1993.pdf

Skinner, E. A., Wellborn, J. G., & Connell, J. P. (1990). What it takes to do well in school and whether I've got it: The role of perceived control in children's engagement and school achievement. *Journal of Educational Psychology, 82*(1), 22–32.

Soodak, L. C., & Podell, P. M. (1996). Teacher efficacy: Toward the understanding of a multi-faceted construct. *Teaching and Teacher Education, 12*(4), 401–411. doi:10.1016/0742-051X(95)00047-N

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*, 214-221.

Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimate approach. *Multivariate Behavioral Research, 25*(2), 173-180. doi:10.1207/s15327906mbr2502_4

Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Suhr, D., D. (2006). *Exploratory or confirmatory factor analysis? Proceedings of the Thirty first Annual SAS® Users Group International Conference*, paper 200-31, San Francisco, CA.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Boston: Pearson Education.

Tanaka, J. S., & Huba, G. J. (1989). A general coefficient of determination for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology, 42,* 233-239.

Tracz, S. M., & Gibson, S. (1986). *Effects of efficacy on academic achievement*. Paper presented at the Annual Meeting of the California Educational Research Association, Marina del Rey, CA.

Tschannen-Moran, M. (n.d.). Research Tools - Megan Tschannen-Moran's Web Site. Retrieved June 7, 2015, from http://wmpeople.wm.edu/index.php/site/page/mxtsch/researchtools

Tschannen-Moran, M., & McMaster, P. (2009). Sources of self efficacy: Four professional development formats and their relationship to self efficacy and implementation of a new teaching strategy. *The Elementary School Journal,110*(2), 246. Retrieved from http://0-search.proquest.com.library.uark.edu/docview/224516793?accountid=8361

Tschannen-Moran, M., & Woolfolk Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*(7), 783-805. doi:10.1016/S0742-051X(01)00036-1

Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*(2), 202-248. doi:10.2307/1170754

Tsigilis, N., Grammatikopoulos, V., & Koustelios, A. (2007). Applicability of the Teachers' Sense of Efficacy Scale to educators teaching innovative programs. *International Journal of Educational Management, 21*(7), 634 – 642.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1-10. doi:10.1007/BF02291170

Turner, R. C., & Carlson, L. (2003). Indexes of item-objective congruence for multidimensional items. *International Journal of Testing, 3*(2), 163.

Tyrer, S., Wigham, A., Cicchetti, D., Margallo-Lana, M., Moore, P., & Reid, B. (2010). Comparison of short and long versions of the Prudhoe cognitive function test and the K-BIT in participants with intellectual impairment. *Journal of Autism & Developmental Disorders, 40*(8), 1000-1005. doi:10.1007/s10803-010-0949-3

Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology*, *34*(1), 89-101.

Vasquez, A. (2009). Teacher efficacy and student achievement in ninth and tenth grade reading: A multilevel analysis. *Dissertation Abstracts International Section A, 70,* 525.

Wang, Y, Yao, L., Liu, L., Yang, X., Wu, H., Wang, J. & Wang, L. (2014). The mediating role of self-efficacy in the relationship between Big five personality and depressive symptoms

among Chinese unemployed population: A cross-sectional study. *BMC Psychiatry,14*(61), 1-8. doi:10.1186/1471-244X-14-61.

Wasserman, J. D., & Bracken, B. A. (2003). Psychometric characteristics of assessment procedures. In I. B. Weiner (Ed.), *Handbook of psychology* (Vol. 10). Hoboken, NJ: John Wiley & Sons.

Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology, 71*(1), 3-25. doi:10.1037/0022-0663.71.1.3

Wolters, C. A., & Daugherty, S. G. (2007). Goals structures and teachers' sense of efficacy: Their relation and association to teaching experience and academic level. *Journal of Educational Psychology, 99* (1), 181-193.

Woolfolk Hoy, A. (2000, April). *Changes in teacher efficacy during the early years of teaching*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Woolfolk Hoy, A., & Burke-Spero, R. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education, 21*(4), 343-356. doi:10.1016/j.tate.2005.01.007

Yousuf Zai, S. A., & Munshi, P. (2016). Construct validity to examine the latent traits of teacher self-efficacy instrument. *Advances in Social Sciences Research Journal, 3*(4), 74-83. DOI: 10.14738/assrj.34.1944.

Yousuf Zai, S. A., & Soomro, K. A., (2015, April). *Examining the latent structure of the teachers' sense of efficacy scale*. Paper presented at the annual meeting of the American Educational Research Association, Chicago Illinois.

Table 5

Demographics of In-Service Teachers (N = 549)

Variable	Categories	Frequency	Percentage
Age group			
	<25	68	12.39%
	25 to 29	109	19.85%
	30 to 34	110	20.04%
	35 to 39	69	12.57%
	40 to 44	90	16.39%
	44+	103	18.76%
Gender			
	Male	171	31.15%
	Female	378	68.85%
Grade Level			
	Primary	119	21.76%
	Elementary	150	27.42%
	Secondary	179	32.72%
	Higher Secondary	71	12.98%
	College/post-graduate	28	5.12%
Academic			
	Bachelor	146	26.59%
	Master	366	66.67%
	M.S. / M. Phil	37	6.74%
Professional			
	B. Ed	244	44.53%
	M. Ed	272	49.64%
	Other (CT/PT)	14	2.55%
	None	18	3.28%
Job Status			
	Permanent	448	81.60%
	Temporary	101	18.40%

Table 6

Demographics of Pre-service Teachers (N = 423)

Variable	Categories	Frequency	Percentage
Age group			
	<25	316	74.70%
	25 to 29	77	18.20%
	30 to 34	20	4.73%
	35 to 39	5	1.18%
	40 to 44	5	1.18%
Gender			
	Male	116	27.42%
	Female	307	72.58%
Academic			
	High School	197	46.57%
	Bachelor	117	27.66%
	Master	98	23.17%
	M.S. / M. Phil	11	2.60%

Table 7

The Index of the Item-Objective Congruence of 24-item TSES

Item	IIOC	Mean rating on valid objective	Mean rating on item objective
1. to get through to the most difficult students	0.79	0.71	-0.33
2. to help your students think critically	0.86	1.00	-0.14
3. to control disruptive behavior in the classroom	1.00	1.00	-0.33
4. to motivate students who show low interest	1.00	1.00	-0.33
5. your expectations clear about student behavior	0.79	0.71	-0.33
6. to get students to believe they can do well in school work	0.79	1.00	-0.05
7. to difficult questions from your students	0.79	1.00	-0.05
8. establish routines to keep activities running smoothly	0.96	1.00	-0.29
9. to help your students value learning	0.75	1.00	0.00
10. gauge student comprehension of what you have taught	1.00	1.00	-0.33
11. craft good questions for your students	1.00	1.00	-0.33
12. to foster student creativity	0.93	1.00	-0.24
13. to get children to follow classroom rules	1.00	1.00	-0.33
14. to improve the understanding of a student	0.82	1.00	-0.10
15. to calm a student who is disruptive or noisy	1.00	1.00	-0.33
16. establish a classroom management system	1.00	1.00	-0.33
17. to adjust lessons to the proper level	1.00	1.00	-0.33
18. variety of assessment strategies	1.00	1.00	-0.33
19. keep a few problem students form ruining an entire lesson	1.00	1.00	-0.33
20. provide an alternative explanation	1.00	1.00	-0.33
21. respond to defiant students	1.00	1.00	-0.33
22. assist families in helping their children do well in school	0.68	0.57	-0.33
23. implement alternative strategies in your classroom	0.79	0.71	-0.33
24. appropriate challenges for very capable students	0.89	1.00	-0.19

Table 8

Mean, Standard Deviation, Skewness, and Kurtosis of the Short and Long Forms TSES for both Pre-service and In-service Teachers

	Short-Form TSES (12-item)				Long-Form TSES (24-item)			
	M	*SD*	Skew	Kurtosis	*M*	*SD*	Skew	Kurtosis
In-service Teachers (*N* = 549)								
Student Engagement	7.53	0.92	-0.41	-0.41	7.49	0.85	-0.34	-0.56
Instructional Strategies	7.48	1.02	-0.63	-0.04	7.49	0.94	-0.64	-0.14
Classroom Management	7.67	1.08	-0.73	0.05	7.54	0.99	-0.63	-0.09
TSES	7.56	0.83	-0.54	-0.05	7.51	0.79	-0.50	-0.10
Pre-service Teachers (*N* = 423)								
Student Engagement	7.08	1.29	-0.78	0.58	6.81	1.27	-0.59	0.09
Instructional Strategies	6.87	1.37	-0.72	0.24	6.89	1.25	-0.74	0.21
Classroom Management	7.07	1.36	-0.92	1.03	6.95	1.21	-0.68	0.11
TSES	7.01	1.05	-0.77	0.56	6.88	1.03	-0.71	0.27

Table 9

Reliability Coefficients of the Participants for both Short-form and Long-form TSES

		Short-Form TSES				Long-Form TSES			
	N	TSES	SE	IS	CM	TSES	SE	IS	CM
Pre-service Teachers	423	0.82	0.70	0.75	0.73	0.91	0.83	0.85	0.82
In-service Teachers	549	0.86	0.69	0.75	0.80	0.92	0.82	0.85	0.86
All	972	0.84	0.70	0.73	0.70	0.92	0.83	0.84	0.83

Table 10

Inter-scale Correlations among Subscales of the TSES for In-service and Pre-service Teachers

	In-service Teachers				Pre-service Teachers			
	TSES	SE	IS	CM	TSES	SE	IS	CM
TSES		0.82*	0.83*	0.82*		0.77*	0.80*	0.78*
Student Engagement	0.84*		0.55	0.51*	0.82*		0.43*	0.39*
Instructional Strategies	0.86*	0.62*		0.49*	0.82*	0.49*		0.43*
Classroom Management	0.85*	0.57*	0.58*		0.84*	0.56*	0.55*	

Note: Above diagonal are short-form (12-item); below diagonal are long-form (24-item).
* $p < .001$.

Table 11

Correlation Matrix for 12-item TSES Scale with Mean and Standard Deviation for In-service Teachers (N = 549)

	2	3	4	11	5	9	10	12	1	6	7	8
M	7.58	7.46	7.68	7.41	7.48	7.31	7.71	7.43	7.69	7.71	7.70	7.58
SD	1.33	1.29	1.31	1.23	1.31	1.41	1.36	1.34	1.33	1.30	1.42	1.38
TSES2	1.00											
TSES3	0.41	1.00										
TSES4	0.36	0.43	1.00									
TSES11	0.29	0.31	0.29	1.00								
TSES5	0.29	0.32	0.35	0.28	1.00							
TSES9	0.33	0.25	0.33	0.26	0.45	1.00						
TSES10	0.34	0.28	0.29	0.22	0.45	0.38	1.00					
TSES12	0.35	0.27	0.29	0.29	0.41	0.47	0.40	1.00				
TSES1	0.31	0.23	0.33	0.22	0.30	0.23	0.28	0.30	1.00			
TSES6	0.32	0.23	0.26	0.22	0.23	0.28	0.33	0.30	0.54	1.00		
TSES7	0.33	0.32	0.35	0.25	0.22	0.29	0.22	0.40	0.49	0.47	1.00	
TSES8	0.34	0.30	0.34	0.25	0.29	0.27	0.33	0.44	0.56	0.48	0.51	1.00

Table 12

Correlation Matrix for 24-item TSES Scale with Mean and Standard Deviation for In-service Teachers (N= 549)

	1	2	4	6	9	12	14	22	7	10	11	17	18	20	23	24	3	5	8	13	15	16	19	21
M	7.49	7.37	7.58	7.46	7.68	7.45	7.49	7.41	7.43	7.60	7.48	7.55	7.31	7.71	7.43	7.42	7.69	7.31	7.43	7.71	7.70	7.58	7.48	7.44
SD	1.25	1.28	1.33	1.29	1.31	1.36	1.15	1.23	1.45	1.29	1.31	1.23	1.41	1.36	1.34	1.40	1.33	1.48	1.38	1.30	1.42	1.38	1.48	1.37
1	1.00																							
2	0.40	1.00																						
4	0.37	0.42	1.00																					
6	0.44	0.48	0.41	1.00																				
9	0.49	0.40	0.36	0.43	1.00																			
12	0.31	0.32	0.39	0.37	0.37	1.00																		
14	0.37	0.44	0.31	0.37	0.30	0.27	1.00																	
22	0.34	0.29	0.29	0.31	0.29	0.24	0.27	1.00																
7	0.31	0.28	0.26	0.28	0.28	0.25	0.28	0.23	1.00															
10	0.34	0.28	0.24	0.29	0.33	0.24	0.30	0.30	0.40	1.00														
11	0.25	0.32	0.29	0.32	0.35	0.28	0.19	0.28	0.45	0.42	1.00													
17	0.35	0.34	0.33	0.25	0.29	0.24	0.34	0.23	0.37	0.41	0.40	1.00												
18	0.30	0.28	0.33	0.25	0.33	0.35	0.29	0.26	0.35	0.39	0.45	0.44	1.00											
20	0.29	0.28	0.34	0.28	0.29	0.35	0.25	0.22	0.34	0.41	0.45	0.36	0.38	1.00										
23	0.35	0.34	0.35	0.27	0.29	0.31	0.27	0.29	0.40	0.46	0.41	0.54	0.47	0.40	1.00									
24	0.28	0.33	0.24	0.30	0.26	0.27	0.23	0.30	0.35	0.39	0.42	0.49	0.39	0.38	0.51	1.00								
3	0.30	0.25	0.31	0.23	0.33	0.26	0.23	0.22	0.29	0.31	0.30	0.30	0.23	0.28	0.30	0.19	1.00							
5	0.33	0.25	0.25	0.28	0.30	0.19	0.26	0.11	0.25	0.26	0.22	0.34	0.18	0.19	0.25	0.26	0.42	1.00						
8	0.34	0.28	0.22	0.26	0.31	0.25	0.23	0.21	0.37	0.28	0.27	0.30	0.29	0.27	0.30	0.32	0.38	0.42	1.00					
13	0.31	0.26	0.32	0.23	0.26	0.29	0.25	0.22	0.30	0.26	0.23	0.30	0.28	0.33	0.30	0.25	0.54	0.39	0.36	1.00				
15	0.36	0.31	0.33	0.32	0.35	0.21	0.33	0.25	0.33	0.35	0.22	0.41	0.29	0.22	0.40	0.31	0.49	0.38	0.33	0.47	1.00			
16	0.39	0.34	0.34	0.30	0.34	0.26	0.33	0.25	0.30	0.36	0.29	0.40	0.27	0.33	0.44	0.30	0.56	0.48	0.37	0.48	0.51	1.00		
19	0.31	0.23	0.30	0.28	0.30	0.25	0.25	0.27	0.26	0.25	0.31	0.32	0.24	0.25	0.33	0.27	0.57	0.43	0.35	0.37	0.50	0.55	1.00	
21	0.22	0.25	0.18	0.22	0.18	0.16	0.23	0.30	0.29	0.25	0.33	0.28	0.26	0.27	0.32	0.30	0.39	0.36	0.39	0.34	0.35	0.45	0.46	1.00

Table 13

Correlation Matrix for 12-item TSES Scale with Mean and Standard Deviation for Pre-service Teachers (N = 423)

	2	3	4	11	5	9	10	12	1	6	7	8
M	7.18	7.17	7.15	6.81	6.91	6.83	7.00	6.72	6.87	7.25	7.08	7.09
SD	1.75	1.77	1.66	1.89	1.74	1.82	1.87	1.82	1.95	1.81	1.77	1.77
TSES2	1.00											
TSES3	0.38	1.00										
TSES4	0.41	0.40	1.00									
TSES11	0.32	0.43	0.30	1.00								
TSES5	0.27	0.26	0.29	0.23	1.00							
TSES9	0.23	0.21	0.23	0.18	0.36	1.00						
TSES10	0.24	0.26	0.31	0.25	0.49	0.42	1.00					
TSES12	0.25	0.15	0.27	0.20	0.41	0.40	0.50	1.00				
TSES1	0.17	0.30	0.15	0.18	0.32	0.26	0.21	0.19	1.00			
TSES6	0.25	0.19	0.21	0.12	0.21	0.28	0.21	0.22	0.42	1.00		
TSES7	0.24	0.26	0.17	0.22	0.31	0.18	0.23	0.13	0.39	0.36	1.00	
TSES8	0.27	0.21	0.19	0.24	0.29	0.25	0.29	0.34	0.44	0.49	0.32	1.00

Table 14

Correlation Matrix for 24-item TSES Scale with Mean and Standard Deviation for Pre-service Teachers (N= 423)

	1	2	4	6	9	12	14	22	7	10	11	17	18	20	23	24	3	5	8	13	15	16	19	21
M	6.45	6.56	7.18	7.17	7.15	6.74	6.45	6.81	6.90	6.89	6.91	6.99	6.83	7.00	6.72	6.87	6.87	6.60	6.87	7.25	7.08	7.09	6.93	6.88
SD	2.06	1.94	1.75	1.77	1.66	1.77	2.19	1.89	1.81	1.81	1.74	1.74	1.82	1.87	1.82	1.81	1.95	1.75	1.71	1.81	1.77	1.77	1.84	1.85
1	1.00																							
2	0.40	1.00																						
4	0.44	0.42	1.00																					
6	0.34	0.34	0.38	1.00																				
9	0.32	0.30	0.41	0.40	1.00																			
12	0.32	0.39	0.39	0.50	0.42	1.00																		
14	0.33	0.33	0.32	0.45	0.40	0.56	1.00																	
22	0.29	0.35	0.32	0.43	0.30	0.33	0.42	1.00																
7	0.19	0.18	0.21	0.29	0.24	0.19	0.24	0.20	1.00															
10	0.24	0.18	0.23	0.22	0.27	0.23	0.23	0.18	0.32	1.00														
11	0.20	0.20	0.27	0.26	0.29	0.22	0.30	0.23	0.37	0.41	1.00													
17	0.21	0.14	0.26	0.20	0.20	0.21	0.22	0.22	0.39	0.47	0.44	1.00												
18	0.21	0.16	0.23	0.21	0.23	0.25	0.26	0.18	0.35	0.42	0.36	0.36	1.00											
20	0.24	0.24	0.24	0.26	0.31	0.28	0.29	0.25	0.45	0.37	0.49	0.56	0.42	1.00										
23	0.22	0.13	0.25	0.15	0.27	0.21	0.23	0.20	0.31	0.43	0.41	0.55	0.40	0.50	1.00									
24	0.30	0.17	0.21	0.23	0.26	0.26	0.28	0.27	0.36	0.38	0.34	0.36	0.45	0.38	0.32	1.00								
3	0.25	0.22	0.17	0.30	0.15	0.22	0.28	0.18	0.32	0.22	0.32	0.17	0.26	0.21	0.19	0.19	1.00							
5	0.24	0.25	0.30	0.29	0.22	0.17	0.26	0.28	0.29	0.27	0.29	0.29	0.25	0.33	0.29	0.30	0.34	1.00						
8	0.33	0.25	0.29	0.33	0.30	0.27	0.30	0.27	0.31	0.30	0.27	0.21	0.28	0.22	0.33	0.32	0.33	0.43	1.00					
13	0.24	0.22	0.25	0.19	0.21	0.19	0.23	0.12	0.20	0.21	0.21	0.20	0.28	0.21	0.22	0.16	0.42	0.31	0.38	1.00				
15	0.29	0.26	0.24	0.26	0.17	0.21	0.35	0.22	0.20	0.30	0.31	0.15	0.18	0.23	0.13	0.18	0.39	0.39	0.32	0.36	1.00			
16	0.33	0.23	0.27	0.21	0.19	0.27	0.30	0.24	0.29	0.28	0.29	0.32	0.25	0.29	0.34	0.24	0.44	0.34	0.32	0.49	0.32	1.00		
19	0.35	0.19	0.29	0.29	0.26	0.24	0.28	0.24	0.31	0.28	0.30	0.22	0.27	0.29	0.25	0.29	0.39	0.33	0.41	0.34	0.43	0.37	1.00	
21	0.29	0.17	0.32	0.25	0.32	0.20	0.36	0.20	0.28	0.23	0.30	0.27	0.19	0.32	0.30	0.29	0.31	0.32	0.29	0.33	0.42	0.45	0.39	1.00

Table 15

Factor Loadings of the CFA Model – I (24-item TSES)

Latent Trait	Item	Unstandardized Loadings [b]	*t* value	Standardized Loadings	Error	R Squared
Student Engagement						
	TSES1	1.00 [a]		0.66*	0.75	0.44
	TSES2	1.01 (0.08)	13.09	0.66*	0.76	0.43
	TSES4	0.98 (0.08)	12.35	0.61*	0.79	0.37
	TSES6	1.03 (0.08)	13.18	0.66*	0.75	0.44
	TSES9	1.02 (0.08)	12.88	0.64*	0.77	0.41
	TSES12	0.89 (0.08)	11.07	0.54*	0.84	0.29
	TSES14	0.76 (0.07)	11.43	0.56*	0.83	0.31
	TSES22	0.72 (0.07)	10.05	0.49*	0.87	0.24
Instructional Strategies						
	TSES7	1.00 [a]		0.59*	0.81	0.35
	TSES10	0.97 (0.08)	11.82	0.64*	0.77	0.41
	TSES11	1.00 (0.08)	11.93	0.65*	0.76	0.42
	TSES17	0.99 (0.08)	12.36	0.68*	0.73	0.46
	TSES18	1.05 (0.09)	11.75	0.63*	0.77	0.40
	TSES20	0.96 (0.09)	11.27	0.60*	0.80	0.36
	TSES23	1.14 (0.09)	12.81	0.72*	0.69	0.52
	TSES24	1.06 (0.09)	11.90	0.65*	0.76	0.42
Classroom Management						
	TSES3	1.00[a]		0.73*	0.68	0.53
	TSES5	0.93 (0.07)	13.46	0.61*	0.79	0.37
	TSES8	0.79 (0.06)	12.29	0.56*	0.83	0.31
	TSES13	0.86 (0.06)	14.20	0.64*	0.77	0.41
	TSES15	0.93 (0.07)	14.98	0.68*	0.74	0.46
	TSES16	1.09 (0.06)	16.92	0.76*	0.65	0.58
	TSES19	1.08 (0.07)	15.75	0.71*	0.70	0.47
	TSES21	0.83 (0.06)	12.88	0.58*	0.81	0.34

[a] First item of each latent factor was set to one.

[b] Standard Error (SE) of unstandardized loadings are presented in bracket.

* Significant at $p < .001$

Table 16

Factor Loadings of the CFA Model – II (12-item TSES)

Item # in short and long-form [c]	Unstandardized Loadings [b]	*t* value	Standardized Loadings	Error Variance	R squared
Student Engagement					
t2(TSES4)	1.00[a]		0.64*	0.68	0.54
t3(TSES6)	0.95(0.09)	11.03	0.62*	0.79	0.38
t4(TSES9)	0.98(0.09)	11.16	0.63*	0.77	0.40
t11(TSES12)	1.05(0.09)	12.39	0.49*	0.87	0.24
Instructional Strategies					
t5(TSES11)	1.00[a]		0.66*	0.75	0.43
t9(TSES18)	1.07(0.09)	12.19	0.66*	0.76	0.43
t10(TSES20)	0.99(0.08)	11.74	0.63*	0.78	0.39
t12(TSES23)	1.05(0.09)	12.39	0.68*	0.74	0.46
Classroom Management					
t1(TSES3)	1.00[a]		0.74*	0.68	0.54
t6(TSES13)	0.9(0.06)	14.32	0.68*	0.73	0.46
t7(TSES15)	0.99(0.07)	14.35	0.68*	0.73	0.47
t8(TSES16)	1.06(0.07)	15.51	0.75*	0.66	0.56

[a] First item of each latent factor was set to one.
[b] Standard Error (SE) of unstandardized loadings are presented in bracket.
[c] Item numbers of short-form TSES appeared with suffix of letter 'b' and item numbers within brackets are appeared in the long-form TSES.
* Significant at $p < .001$

Table 17

Factor Loadings (Standardized Regression Coefficients) for a Three-Factor of the 24-item TSES for Pre-service Teachers (N = 423)

Items	Factor1	Factor2	Factor3
Student Engagement			
TSES1		.41*	
TSES2		.55*	
TSES4		.52*	
TSES6		.68*	
TSES9		.56*	
TSES12		.76*	
TSES14		.62*	
TSES22		.54*	
Instructional Strategies			
TSES7	.45*		
TSES10	.56*		
TSES11	.54*		
TSES17	.82*		
TSES18	.52*		
TSES20	.72*		
TSES23	.73*		
TSES24	.46*		
Classroom Management			
TSES3			.67*
TSES5			.44*
TSES8			.44*
TSES13			.67*
TSES15			.63*
TSES16			.58*
TSES19			.55*
TSES21			.47*

Note: Values greater than .35 are flagged by '*'

Table 18

Factor Loadings (Standardized Regression Coefficients) for a Three-Factor of the 12-item TSES for Pre-service Teachers (N = 423)

Items	Factor1	Factor2	Factor3
Student Engagement			
TSES2			.66*
TSES3			.79*
TSES4			.67*
TSES11			.72*
Instructional Strategies			
TSES5	.62*		
TSES9	.68*		
TSES10	.79*		
TSES12	.84*		
Classroom Management			
TSES3		.78*	
TSES5		.78*	
TSES8		.66*	
TSES21		.68*	

Note: Values greater than .43 are flagged by '*'

Table 19

Multitrait-Multimethod (MTMM) Raw Correlation Matrix, Means (M) and Standard Deviation (SD) for Three Subscales of the TSES and Three Alternative Measures (N = 549)

Method		Method 1 (TSES)			Method 2 (Alternative Measures)		
	Scales	SE	IS	CM	IMS	PTE	BMS
TSES							
	SE	*(0.82)*					
	IS	0.62	*(0.85)*				
	CM	0.57	0.58	*(0.86)*			
Alternative	Measures						
	IMS	**0.42**	0.37	0.32	*(0.80)*		
	PTE	0.41	**0.43**	0.36	0.49	*(0.85)*	
	BMS	0.40	0.39	**0.41**	0.36	0.40	*(0.81)*
	M	7.49	7.49	7.54	4.95	4.96	4.38
	SD	0.85	0.94	0.99	0.66	0.58	0.50

Note: SE = Student Engagement; IS = Instructional Strategies; CM = Classroom Management; IMS = Instructional Management Scale; PTE = Personal Teaching Efficacy; BMS = Behavior Management Scale. Reliability coefficients of each scale presented on the main diagonal in italicized text within parentheses; Convergent validities (Monotrait-heteromethod) appear as bolded font; Correlations for different traits using the same method (Heteromethod-monotrait correlation) are underlined; the remaining correlations are from different scales and using different methods (Heterotrait-heteromethod correlations).

Table 20

Multitrait Multimethod Correlation (MTMM) Matrix of Scales after Correction for Attenuation (N = 549)

Method		Method 1 (TSES)			Method 2 (Alternative Measures)		
	Scales	SE	IS	CM	IMS	PTE	BMS
TSES							
	SE	*(0.82)*					
	IS	0.75	*(0.85)*				
	CM	0.68	0.69	*(0.86)*			
Alternative Measures							
	IMS	**0.52**	0.45	0.39	*(0.80)*		
	PTE	0.50	**0.50**	0.42	0.59	*(0.85)*	
	BMS	0.49	0.47	**0.49**	0.45	0.48	*(0.81)*

Note: SE = Student Engagement; IS = Instructional Strategies; CM = Classroom Management; IMS = Instructional Management Scale; PTE = Personal Teaching Efficacy; BMS = Behavior Management Scale. Reliability coefficients of each scale presented on the main diagonal in italicized text within parentheses; Convergent validities (Monotrait-heteromethod) appear as bolded font; Correlations for different traits using the same method (Heteromethod-monotrait correlation) are underlined; the remaining correlations are from different scales and using different methods (Heterotrait-heteromethod correlations).

Table 21

Descriptive Statistics of each Subscale's Score of TSES (24-item) for Pre-service Teachers' Gender Group (N = 416)

Level of Gender	*N*	SE		IS		CM	
		M	*SD*	*M*	*SD*	*M*	*SD*
Female	301	6.97	1.14	6.98	1.22	7.09	1.15
Male	115	6.48	1.38	6.77	1.12	6.67	1.15

Table 22

Descriptive Statistics of each Subscale's Score of TSES (24-item) for Pre-service Teachers' Academic Qualification Group (N = 416)

Level of Academic	*N*	SE		IS		CM	
		M	*SD*	*M*	*SD*	*M*	*SD*
High School (XII)	116	6.60	1.26	6.70	1.27	6.61	1.34
Bachelor	193	7.05	1.17	7.03	1.04	7.23	1.01
Master / MS	107	6.71	1.25	6.98	1.34	6.82	1.22

Table 23

Descriptive Statistics of each Score of TSES (24-item) for In-service Teachers' Gender Group (N = 549)

Level of Gender	*N*	SE		IS		CM	
		M	*SD*	*M*	*SD*	*M*	*SD*
Female	378	7.65	0.80	7.66	0.85	7.72	0.92
Male	171	7.14	0.85	7.12	1.02	7.14	1.01

Table 24

Descriptive Statistics of each Subscale's Score of TSES (24-item) for In-service Teachers' at Teaching Grade Level (N = 542)

Teaching Grade Level	*N*	SE		IS		CM	
		M	*SD*	*M*	*SD*	*M*	*SD*
Primary	118	7.85	0.73	7.79	0.81	7.86	0.95
Elementary	148	7.43	0.86	7.51	0.99	7.53	1.03
Secondary	175	7.37	0.82	7.33	0.91	7.40	0.96
Higher Secondary	101	7.41	0.82	7.44	0.90	7.46	0.90

Table 25

Descriptive Statistics of each Subscale's Score of TSES (24-item) for In-service Teachers' Teaching Experience (N = 549)

Teaching Experience group	*N*	SE		IS		CM	
		M	*SD*	*M*	*SD*	*M*	*SD*
2 - 8 years	264	7.43	0.84	7.46	0.93	7.43	0.99
9 – 15 years	110	7.54	0.83	7.33	0.93	7.44	0.93
16 – 22 years	99	7.43	0.86	7.53	0.95	7.85	0.96
More than 22 years	76	7.71	0.84	7.77	0.94	7.67	1.01

Table 26

Descriptive Statistics of each Subscale's Score of TSES (24-item) for In-service Teachers' Job Status (N = 549)

Job Status	*N*	SE		IS		CM	
		M	*SD*	*M*	*SD*	*M*	*SD*
Permanent	448	7.50	0.83	7.51	0.95	7.58	0.97
Temporary (contract)	101	7.45	0.91	7.44	0.89	7.36	1.05

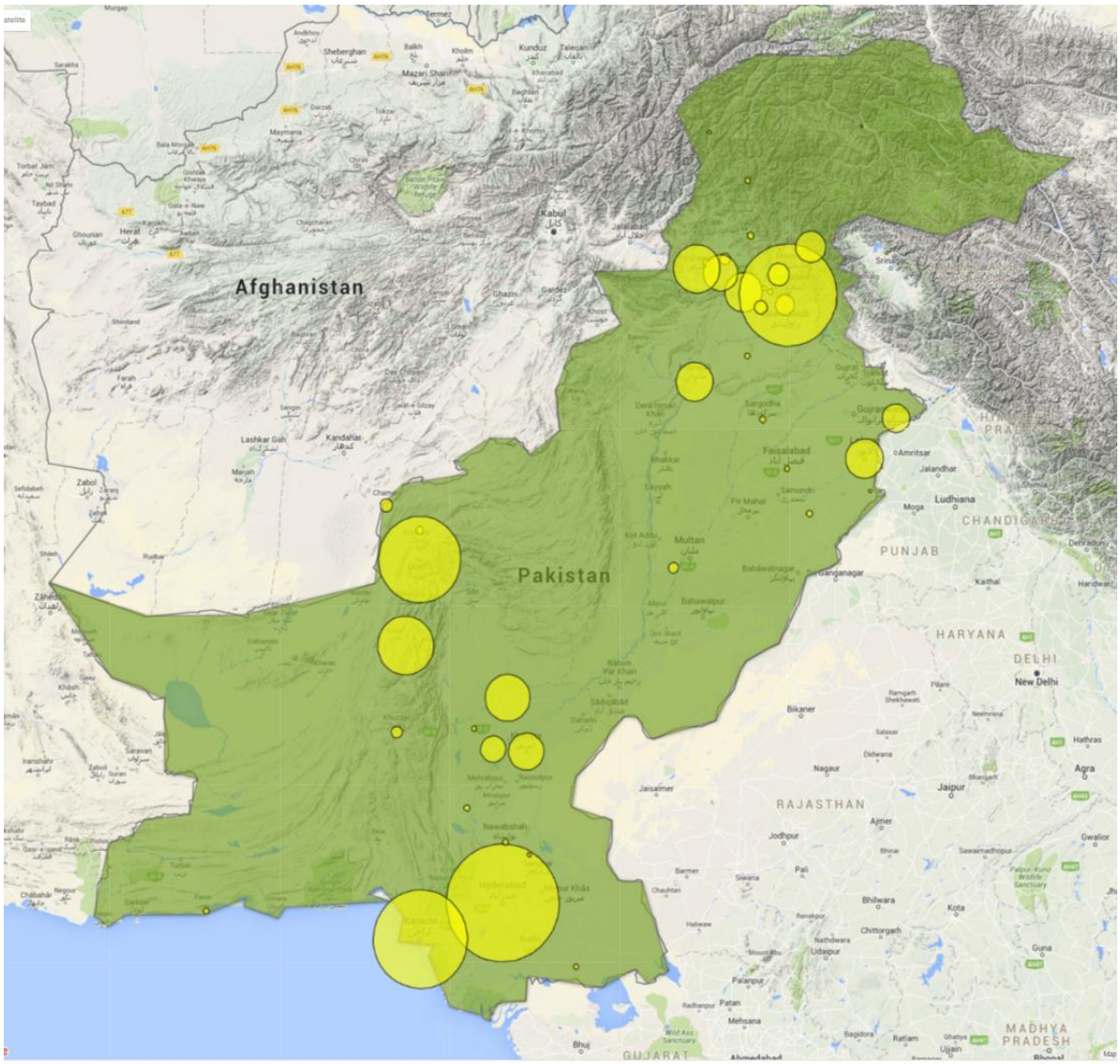

Figure 3. In-service and pre-service teachers' representation from different parts of Pakistan.

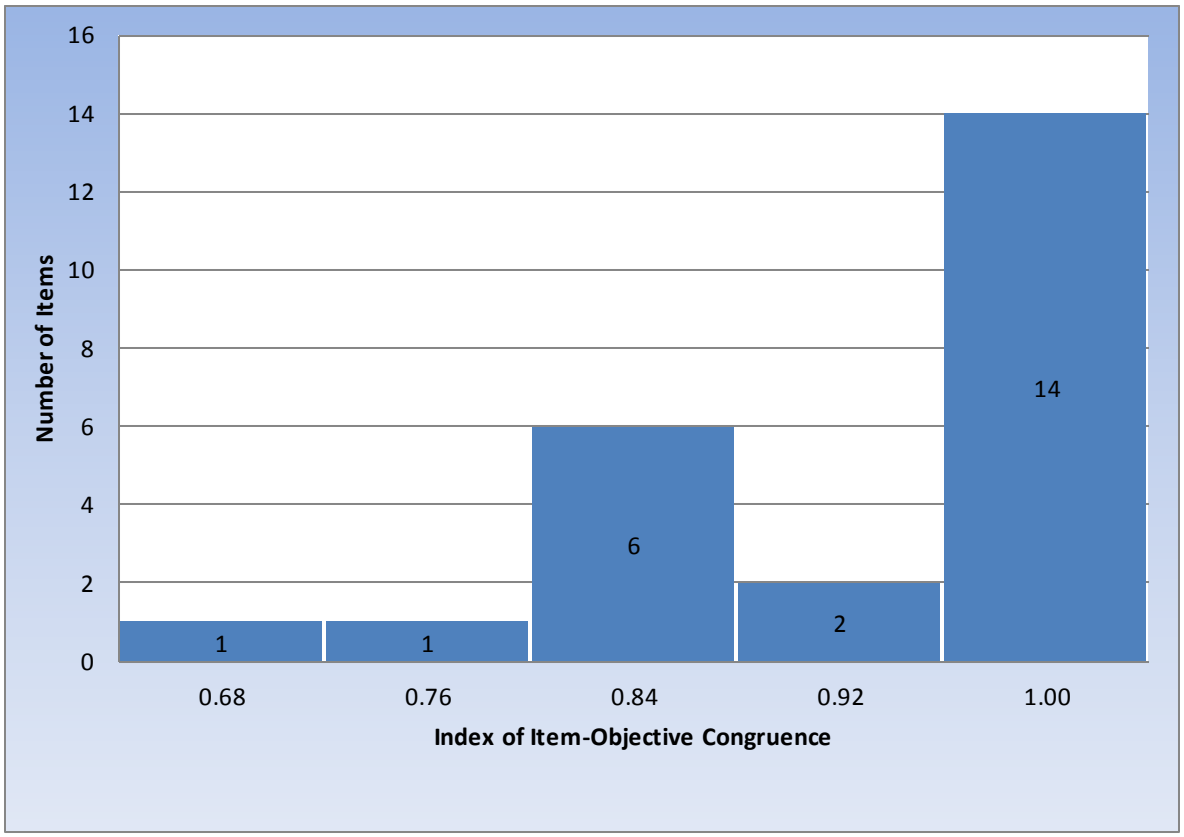

Figure 4. Histogram of item-objective congruence.

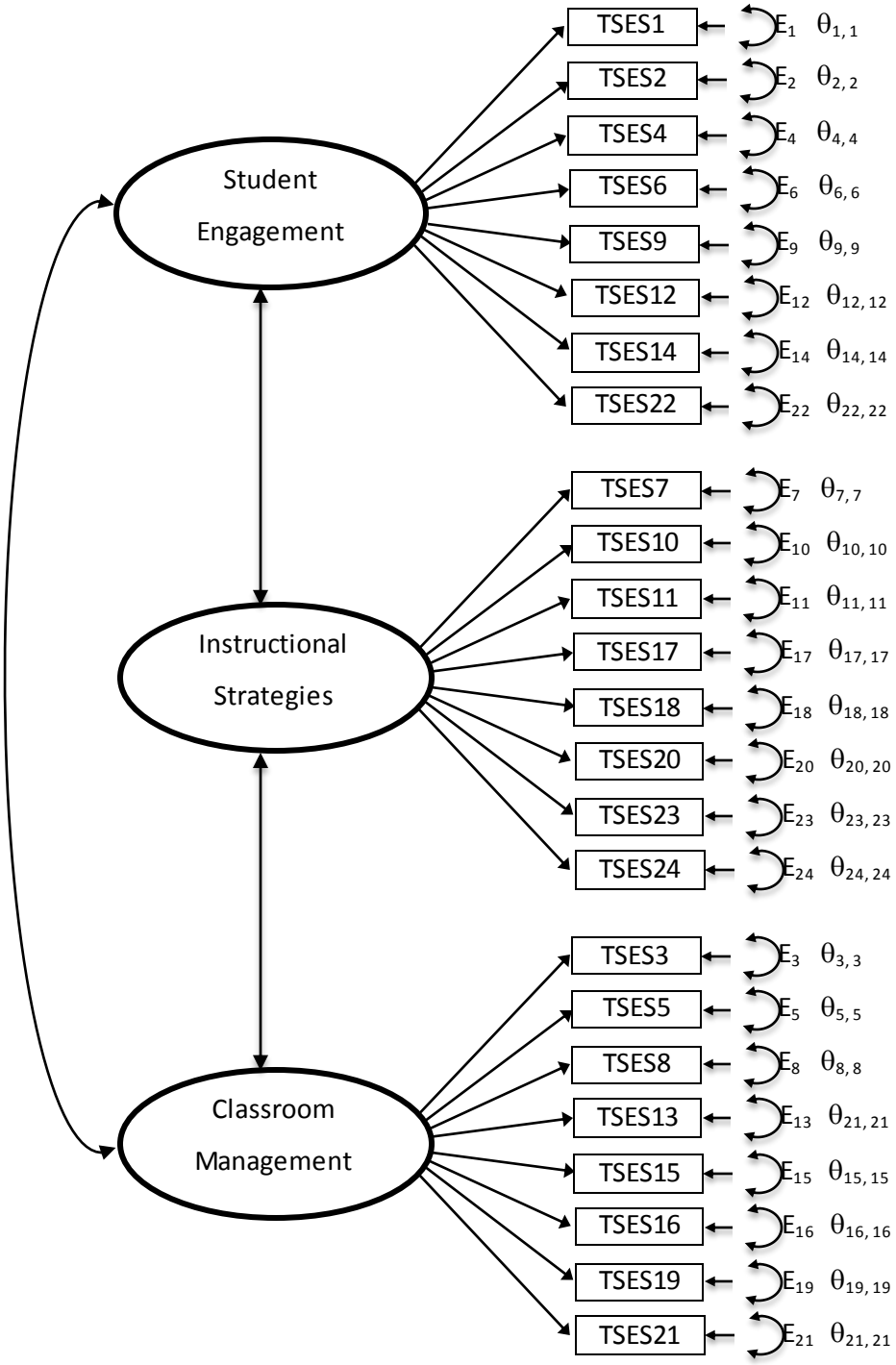

Figure 5. Initial proposed model I for 24-item TSES for in-service teachers.

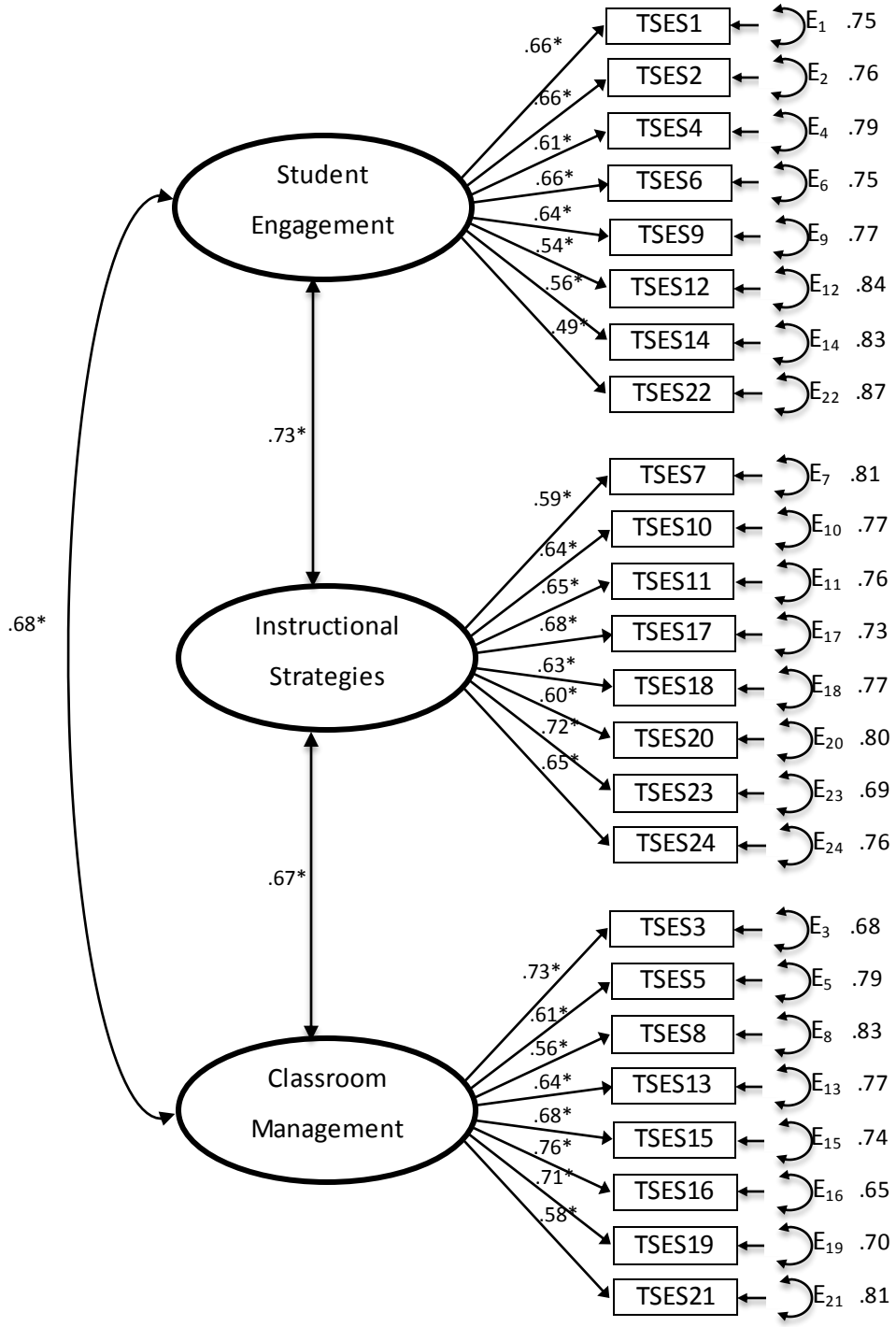

Figure 6. Final model I for the first order correlated three-factor standardized solution (24-item TSES) for in-service teachers.

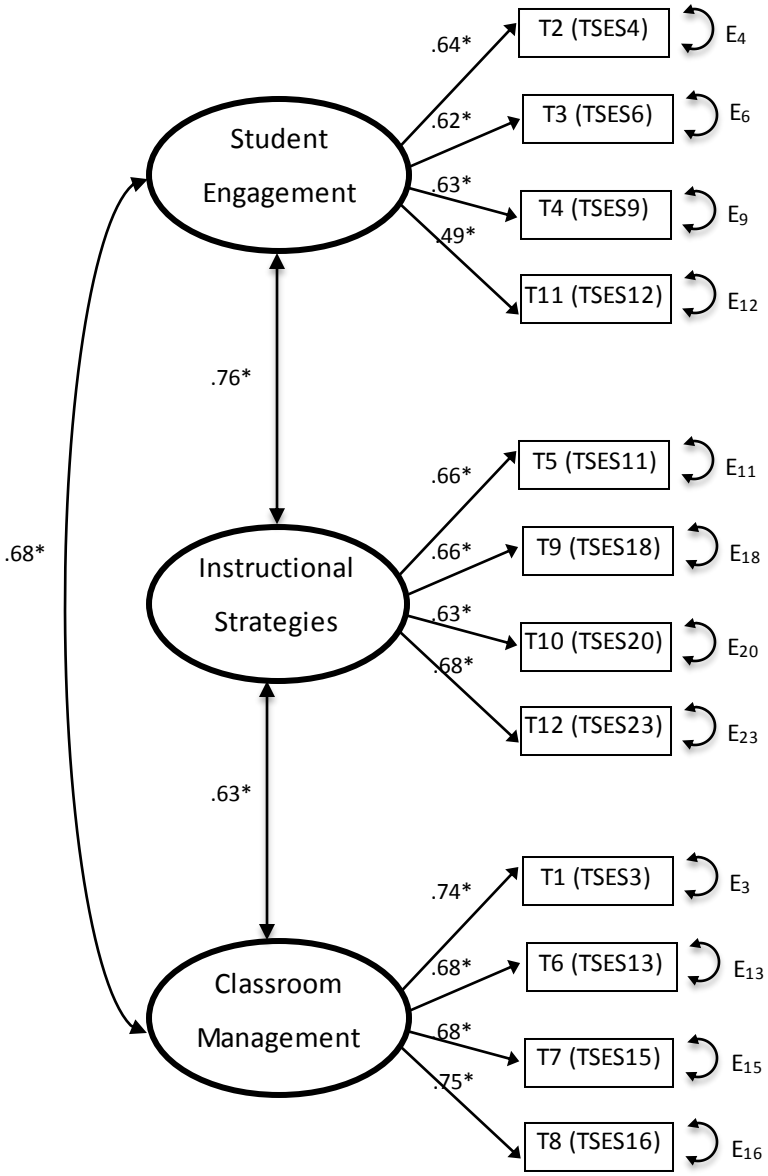

Figure 7. Final model II for the first order correlated three-factor standardized solution (12-item TSES) for in-service teachers.

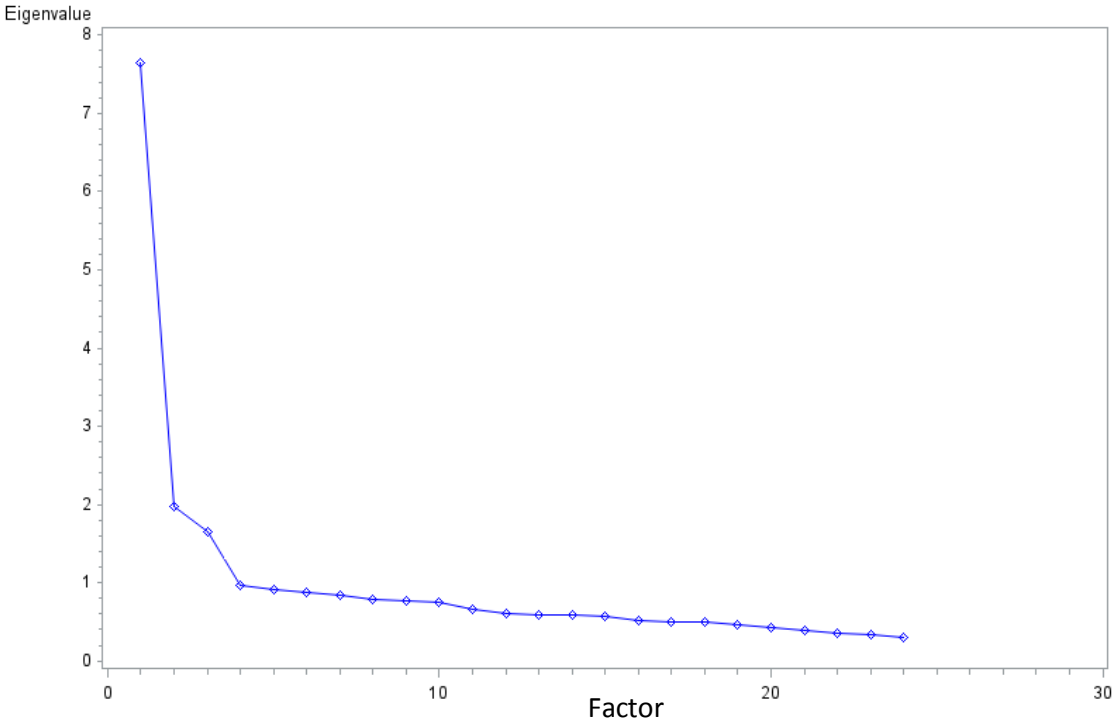

Figure 8. Scree plot of the 24-item TSES using exploratory actor analysis for pre-service teachers.

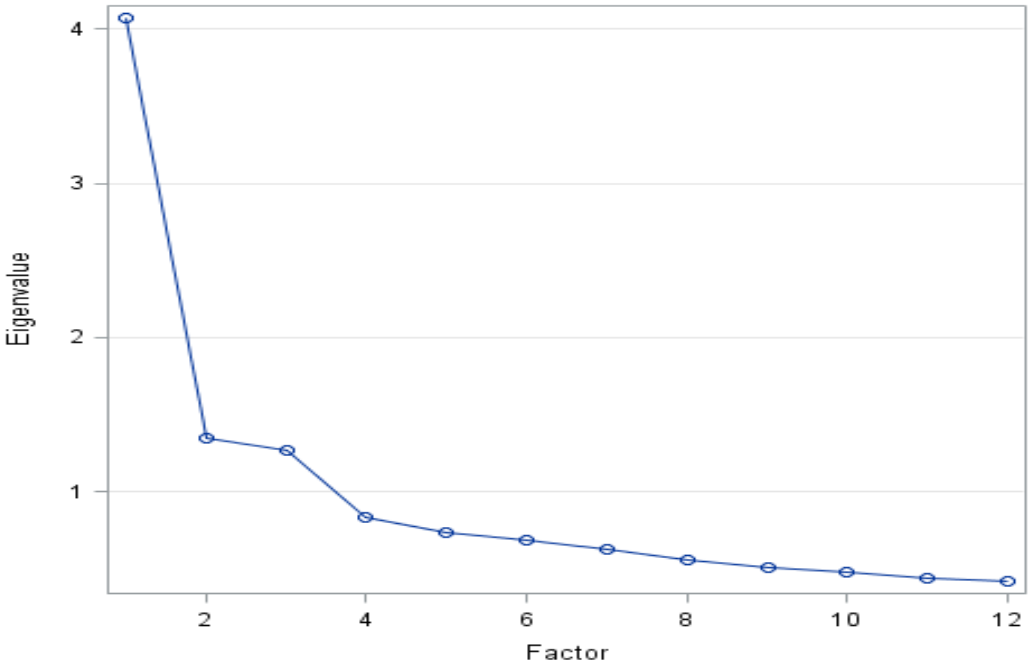

Figure 9. Scree plot of the 12-item TSES using exploratory actor analysis for pre-service teachers.

Appendix A

Teacher Demographic Questionnaire

Please answer the following questions:

1. What is your age group?

   □ Below 25 □ 25 to 29 □ 30 to 34

   □ 35 to 39 □ 40 to 44 □ More than 44

2. How many years have you taught? (please type number of years) _____
3. Please specify your gender?

   □ Male □ Female

4. Which grade level students do you teach?

   □ Primary □ Elementary □ Secondary

   □ Higher Secondary □ College/ University level

5. What is your highest academic qualification?

   □ Bachelor □ Master □ MS / M. Phil □ Ph. D

6. Please specify if you have any professional qualification?

   □ B. Ed □ M. Ed □ Other □ None

7. Please specify your job status?

   □ Permanent □ Temporary (contract)

Appendix B

Teacher' Sense of Efficacy Scale (long-form)

	Teacher Beliefs	How much can you do?								
	Directions: This questionnaire is designed to help us gain a better understanding of the kinds of things that create difficulties for teachers in their school activities. Please circle your opinion about each of the statements below. Your answers are confidential.	**Nothing**		**Very little**		**Some influence**		**Quite a bit**		**A great deal**
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	How much can you do to get through to the most difficult students?	1	2	3	4	5	6	7	8	9
2	How much can you do to help your students think critically?	1	2	3	4	5	6	7	8	9
3	How much can you do to control disruptive behavior in the classroom?	1	2	3	4	5	6	7	8	9
4	How much can you do to motivate students who show low interest in school work?	1	2	3	4	5	6	7	8	9
5	To what extent can you make your expectations clear about student behavior?	1	2	3	4	5	6	7	8	9
6	How much can you do to get students to believe they can do well in school work?	1	2	3	4	5	6	7	8	9
7	How well can you respond to difficult questions from your students?	1	2	3	4	5	6	7	8	9
8	How well can you establish routines to keep activities running smoothly?	1	2	3	4	5	6	7	8	9
9	How much can you do to help your students value learning?	1	2	3	4	5	6	7	8	9
10	How much can you gauge student comprehension of what you have taught?	1	2	3	4	5	6	7	8	9
11	To what extent can you craft good questions for your students?	1	2	3	4	5	6	7	8	9
12	How much can you do to foster student creativity?	1	2	3	4	5	6	7	8	9

	Teacher Beliefs	How much can you do?								
	Directions: This questionnaire is designed to help us gain a better understanding of the kinds of things that create difficulties for teachers in their school activities. Please circle your opinion about each of the statements below. Your answers are confidential.	**Nothing**		**Very little**		**Some influence**		**Quite a bit**		**A great deal**
		1	2	3	4	5	6	7	8	9
13	How much can you do to get children to follow classroom rules?	1	2	3	4	5	6	7	8	9
14	How much can you do to improve the understanding of a student who is failing?	1	2	3	4	5	6	7	8	9
15	How much can you do to calm a student who is disruptive or noisy?	1	2	3	4	5	6	7	8	9
16	How well can you establish a classroom management system with each group of students?	1	2	3	4	5	6	7	8	9
17	How much can you do to adjust your lessons to the proper level for individual students?	1	2	3	4	5	6	7	8	9
18	How much can you use a variety of assessment strategies?	1	2	3	4	5	6	7	8	9
19	How well can you keep a few problem students form ruining an entire lesson?	1	2	3	4	5	6	7	8	9
20	To what extent can you provide an alternative explanation or example when students are confused?	1	2	3	4	5	6	7	8	9
21	How well can you respond to defiant (disobedient) students?	1	2	3	4	5	6	7	8	9
22	How much can you assist families in helping their children do well in school?	1	2	3	4	5	6	7	8	9
23	How well can you implement alternative strategies in your classroom?	1	2	3	4	5	6	7	8	9
24	How well can you provide appropriate challenges for very capable students?	1	2	3	4	5	6	7	8	9

Appendix C

Permission to use the TSES (First author)

William & Mary
School of Education

MEGAN TSCHANNEN-MORAN, PHD
PROFESSOR OF EDUCATIONAL LEADERSHIP

July 9, 2015

Sajid,

You have my permission to use the Teacher Sense of Efficacy Scale (formerly called the Ohio State Teacher Sense of Efficacy Scale), which I developed with Anita Woolfolk Hoy, in your research. You can find a copy of the measure and scoring directions on my web site at http://wmpeople.wm.edu/site/page/mxtsch . Please use the following as the proper citation:

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

I will also attach directions you can follow to access my password protected web site, where you can find the supporting references for this measure as well as other articles I have written on this and related topics.

I would love to receive a brief summary of your results.

All the best,

Megan Tschannen-Moran
The College of William and Mary
School of Education

Appendix D

Permission to use the TSES (Second author)

ANITA WOOLFOLK HOY, PH.D. **PROFESSOR**
PSYCHOLOGICAL STUDIES IN EDUCATION

Dear

You have my permission to use the *Teachers' Sense of Efficacy Scale* in your research. A copy of both the long and short forms of the instrument as well as scoring instructions can be found at:

http://www.coe.ohio-state.edu/ahoy/researchinstruments.htm

Best wishes in your work,

Anita Woolfolk Hoy

Anita Woolfolk Hoy, Ph.D.
Professor

College of Education Phone 614-292-3774
29 West Woodruff Avenue www.coe.ohio-state.edu/ahoy FAX 614-292-7900
Columbus, Ohio 43210-1177
Hoy.17@osu.edu

Appendix E

Behavior Management Strategies Scale (Nie, Lau & Liau, 2012)

Direction: *For each statement below, please mark the response that best describes what you do in the classroom. There are no right or wrong answers, so please respond as honestly as possible.*	**1 Never**	**2 Rarely**	**3 Sometimes**	**4 Frequently**	**5 Always**
I establish specific rules and consequences for student misbehavior					
I monitor the entire classroom					
I correct misbehavior immediately					
I rewards (e.g., praise) good behavior					
I use consistent disciplinary practices					
I discourage misbehavior					
I discuss behavior problem with students to get their prospective.					

Appendix F

Permission to use BIMS scales

Permission to use the sub-scale from the BIMS Inbox x

Sajid Yousuf Zai <sayousuf@email.uark.edu> Jul 26

to Nancy.Martin, daniel.sass

Dear Dr. Martin & Dr. Sass,
I am doctoral candidate in the University of Arkansas and my major is "Educational Statistics and Research Methods". I am writing this email to request your permission to use the ***Instructional Management Scale*** from your published article "Construct Validation of the Behavior and Instructional Management Scale" for my research. My research is based on scale validation procedure and I selected this "***Instructional Management Scale"*** sub-scale to test the construct validity of the TSES scale (Tschannen-Moran & Hoy, 2001).

I hope, I will get positive response from you. If you have any
questions, please feel free to contact me. I look forward to hearing from you soon.

Thank you
--

SAJID A. YOUSUF ZAI

Teaching Assistant, College of Education and Health Professions

Ph.D Scholar, Educational Statistics & Research Methods

University of Arkansas

Fayetteville, AR 72701

(646) 236-0715

Nancy Martin <Nancy.Martin@utsa.edu> Jul 27

to me, Daniel

Yes, you definitely have my permission to use the BIMS in your study. I have attached the article describing its development. A copy of the instrument is included in the appendix at the end of the article.

Good luck with your study.
--

Nancy K. Martin, Ed.D.
Associate Vice Provost — Core Curriculum & QEP
Professor of Educational Psychology
The University of Texas at San Antonio
One UTSA Circle
San Antonio, TX 78249

Appendix G

Personal Teaching Efficacy from Gibson and Dembo (1984) scale

Please indicate the degree to which you agree or disagree with each statement	Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
If a student masters a new math concept quickly, this might be because I knew the necessary steps in teaching that concept.						
When the grades of my students improve it is usually because I found more effective teaching approaches.						
When I really try, I can get through to most difficult students.						
If a student did not remember information I gave in a previous lesson, I would know how to increase his/her retention in the next lesson.						
When a student does better than usual, many times it is because I exerted a little extra effort.						
If a student in my class becomes disruptive and noisy, I feel assured that I know some techniques to redirect him quickly.						
If one of my students could not do a class assignment, I would be able to accurately assess whether the assignment was at the correct level of difficulty.						
When a student is having difficulty with an assignment, I am usually able to adjust it to his/her level						
When a student gets a better grade than he usually gets, it is usually because I found better ways of teaching that student.						

Appendix H

Instructional Management scale from Martin and Sass (2010)

Direction: *For each statement below, please mark the response that best describes what you do in the classroom. There are no right or wrong answers, so please respond as honestly as possible.*	**Strongly disagree**	**Disagree**	**Slightly disagree**	**Slightly agree**	**Agree**	**Strongly agree**
I nearly always use collaborative learning to explore questions in the classroom.						
I engage students in active discussion about issues related to real world applications.						
I nearly always use group work in my classroom.						
I use student input when creating student projects.						
I nearly always adjust instruction in response to individual student needs.						
I nearly always use a teaching approach that encourages interaction among students.						

Thanks for your valuable responses and time

Much Appreciated! ☺

Appendix I

IRB Approval

UNIVERSITY OF ARKANSAS

Office of Research Compliance
Institutional Review Board

November 11, 2015

MEMORANDUM

TO:	Sajid Ali Yousuf Zai Ronna Turner
FROM:	Ro Windwalker IRB Coordinator
RE:	PROJECT MODIFICATION
IRB Protocol #:	15-07-038
Protocol Title:	*Investigating the Factor Structure of the Teachers' Sense of Efficacy Scale (TSES) with Pakistani Teachers*
Review Type:	☒ EXEMPT ☐ EXPEDITED ☐ FULL IRB
Approved Project Period:	Start Date: 11/11/2015 Expiration Date: 08/06/2016

Your request to modify the referenced protocol has been approved by the IRB. **This protocol is currently approved for 1,100 total participants.** If you wish to make any further modifications in the approved protocol, including enrolling more than this number, you must seek approval *prior to* implementing those changes. All modifications should be requested in writing (email is acceptable) and must provide sufficient detail to assess the impact of the change.

Please note that this approval does not extend the Approved Project Period. Should you wish to extend your project beyond the current expiration date, you must submit a request for continuation using the UAF IRB form "Continuing Review for IRB Approved Projects." The request should be sent to the IRB Coordinator, 109 MLKG Building.

For protocols requiring FULL IRB review, please submit your request at least one month prior to the current expiration date. (High-risk protocols may require even more time for approval.) For protocols requiring an EXPEDITED or EXEMPT review, submit your request at least two weeks prior to the current expiration date. Failure to obtain approval for a continuation *on or prior to* the currently approved expiration date will result in termination of the protocol and you will be required to submit a new protocol to the IRB before continuing the project. Data collected past the protocol expiration date may need to be eliminated from the dataset should you wish to publish. Only data collected under a currently approved protocol can be certified by the IRB for any purpose.

If you have questions or need any assistance from the IRB, please contact me at 109 MLKG Building, 5-2208, or irb@uark.edu.

109 MLKG • 1 University of Arkansas • Fayetteville, AR 72701-1201 • (479) 575-2208 • Fax (479) 575-6527 • Email irb@uark.edu

Appendix J

Consent to Participant

Investigating the Factor Structure of the Teachers' Sense of Efficacy Scale (TSES) with Pakistani Teachers

Consent to Participate in a Research Study

Principal Researcher: Sajid Ali Yousuf Zai

Faculty Advisor: Dr. Ronna C. Turner

INVITATION TO PARTICIPATE

You are invited to participate in a research study about investigation the factor structure of the Teachers' Sense of Efficacy Scale (TSES) in Pakistan's context. You are being asked to participate in this study because you are in-service or pre-service public school teachers in Pakistan.

WHAT YOU SHOULD KNOW ABOUT THE RESEARCH STUDY

Who is the Principal Researcher?

Sajid Ali Yousuf Zai, Ph.D scholar of Educational Statistics and Research Methods (ESRM) at the college of Education and Health Professions, University of Arkansas, Fayetteville, AR 72701

[redacted]

sayousuf@uark.edu

Who is the Faculty Advisor?

Ronna C. Turner, Ph.D.
Associate Professor, Educational Statistics & Research Methods
Director, Psychometric and Educational Evaluation Research office
250 Graduate Education Building
University of Arkansas
Fayetteville, AR 72701
(479) 575-2820 (Fax)

What is the purpose of this research study?

The purpose of this study is to investigate the validity of the Teacher's sense of Efficacy for Pakistani in-service and preservice papulation.

Who will participate in this study?

There will be about 1100 Pakistani in-service and pre-service school teachers in this study age ranges from 26 years to 50 years from major cities of Pakistan.

What am I being asked to do?

Your participation will require the following:

There are seven demographic questions and there are 46 questions related to teacher beliefs on three major teaching-related tasks (i.e., classroom management, instructional strategies, and

student engagement). These questions are designed to measure perceived efficacy of preservice and in-service teachers in mentioned three teaching-related tasks.

What are the possible risks or discomforts?
There are no anticipated risks to participating.

What are the possible benefits of this study?
This is scale validation study. The findings will help researchers to know how latent structure of a scale is affected by different culture and education system. The findings will also be useful for teacher's efficacy experts to improve the teacher's efficacy scale world widely.

How long will the study last?
This questionnaire will take approximately 20 minutes to complete.

Will I receive compensation for my time and inconvenience if I choose to participate in this study?
There is no external compensation for participating in the study.

Will I have to pay for anything?
No, there will be no cost associated with your participation.

What are the options if I do not want to be in the study?
If you do not want to be in this study, you may refuse to participate. Also, you may refuse to participate at any time during the study. Your job status, position, scale (grade), and your relationship with the institution, etc. will not be affected in any way if you refuse to participate.

How will my confidentiality be protected?
All information will be kept confidential to the extent allowed by applicable State and Federal law. The Participant's responses will be completely anonymous. Participants will not be asked to write their name or institutional information. Furthermore, all responses in the research will be reported in aggregate form. Therefore, there is no chance of identification of the participants.

Will I know the results of the study?
At the conclusion of the study you will have the right to request feedback about the results. You may contact the faculty advisor, Ronna C. Turner at rcturner@uark.edu or Principal Researcher, Sajid Ali Yousuf Zai at sayousuf@uark.edu. You are encouraged to keep a copy of this email for your files.

What do I do if I have questions about the research study?
You have the right to contact the Principal Researcher or Faculty Advisor as listed below for any concerns that you may have.

Sajid Ali Yousuf Zai
Ph.D scholar, Educational Statistics & Research Methods
College of Education and Health Professions,
University of Arkansas,

Fayetteville, AR 72701
[redacted]
sayousuf@uark.edu

Ronna C. Turner, Ph.D.
Associate Professor, Educational Statistics & Research Methods
Director, Psychometric and Educational Evaluation Research office
250 Graduate Education Building
University of Arkansas
Fayetteville, AR 72701
(479) 575-2820 (Fax)

You may also contact the University of Arkansas Research Compliance office listed below if you have questions about your rights as a participant, or to discuss any concerns about, or problems with the research.

Ro Windwalker, CIP
Institutional Review Board Coordinator
Research Compliance
University of Arkansas
210 Administration
Fayetteville, AR 72701-1201
479-575-2208
irb@uark.edu

I have read the above statement and have been able to ask questions and express concerns, which have been satisfactorily responded to by the investigator. I understand the purpose of the study as well as the potential benefits and risks that are involved. I understand that participation is voluntary. I understand that significant new findings developed during this research will be shared with the participant. I understand that no rights have been waived by signing the consent form. I have been given a copy of the consent form.

_____

Appendix K

EQS Code for Model I

```
/TITLE
  Model I for In-Service Teachers (3-Factors Long-from 24 items
TSES)
/SPECIFICATIONS
Data = 'D:\TSES Analysis\data\InServiceLong.dat';
  VARIABLES=24; CASES=549;
  MATRIX=RAW;   METHOD=ML;
  Analysis = COVARIANCE;
/LABELS
V1=TSES1; V2=TSES2; V3=TSES3; V4=TSES4; V5=TSES5;
V6=TSES6; V7=TSES7; V8=TSES8; V9=TSES9; V10=TSES10;
V11=TSES11; V12=TSES12; V13=TSES13; V14=TSES14; V15=TSES15;
V16=TSES16; V17=TSES17; V18=TSES18; V19=TSES19; V20=TSES20;
V21=TSES21; V22=TSES22; V23=TSES23; V24=TSES24;
F1 = SE; F2= IS; F3=CM;
/EQUATIONS
  V1  = 1F1 + E1;   V2  = *F1 + E2;
  V4  = *F1 + E4;   V6  = *F1 + E6;
  V9  = *F1 + E9;   V12 = *F1 + E12;
  V14 = *F1 + E14;  V22 = *F1 + E22;

    V7  = 1F2 + E7;     V10 = *F2 + E10;
    V11 = *F2 + E11;    V17 = *F2 + E17;
    V18 = *F2 + E18;    V20 = *F2 + E20;
    V23 = *F2 + E23;    V24 = *F2 + E24;
        V3  = 1F3 + E3;     V5  = *F3 + E5;
        V8  = *F3 + E8;     V13 = *F3 + E13;
        V15 = *F3 + E15;    V16 = *F3 + E16;
        V19 = *F3 + E19;    V21= *F3 + E21;
/VARIANCES
E1 to E24=*;
F1 to F3  = *;
/COVARIANCES
F1 to F3 = *;
/LMTEST
SET = PEE, GVF;
/Print
Fit = ALL;
Covariance = Yes;
/END
```

Appendix L

EQS Code for Model II

```
/TITLE
   Model II for In-Service Teachers (3-Factors Short-from 12
items TSES)
/SPECIFICATIONS
Data = 'D:\TSES Analysis\data\InServiceShort.dat';
  VARIABLES=12 ; CASES=550;
  MATRIX=RAW;
  METHOD=ML;
  Analysis = COVARIANCE;
/LABELS
V1= T1; V2=T2; V3=T3; V4=T4; V5=T5;
V6=T6; V7=T7; V8=T8; V9=T9; V10=T10;
V11=T11; V12=T12;
F1 = Se; F2 = IS;  F3= CM;

/EQUATIONS
  V2  = 1F1 + E2;
  V3  = *F1 + E3;
  V4  = *F1 + E4;
  V11 = *F1 + E11;

     V5 = 1F2 + E5;
    V9  = *F2 + E9;
    V10 = *F2 + E10;
    V12 = *F2 + E12;

          V1  = 1F3 + E1;
        V6  = *F3 + E6;
        V7  = *F3 + E7;
        V8 = *F3 + E8;

/VARIANCES
E1 to E12 = *;
/COVARIANCES
F1 to F3 = *;
/LMTEST
SET = PEE, GVF;
/Print
Fit = ALL;
Covariance = Yes;
/END
```

Appendix M

EQS Code for Model III

```
/TITLE
  Model III for Preservice Teachers (1-Factors Long-from 24
items TSES)
/SPECIFICATIONS
Data = 'D:\TSES Analysis\data\PreServiceLong.dat';
  VARIABLES=24 ; CASES=424;
  MATRIX=RAW; METHOD= ML;
  Analysis = COVARIANCE;
/LABELS
V1=TSES1; V2=TSES2; V3=TSES3; V4=TSES4; V5=TSES5;
V6=TSES6; V7=TSES7; V8=TSES8; V9=TSES9; V10=TSES10;
V11=TSES11; V12=TSES12; V13=TSES13; V14=TSES14; V15=TSES15;
V16=TSES16; V17=TSES17; V18=TSES18; V19=TSES19; V20=TSES20;
V21=TSES21; V22=TSES22; V23=TSES23; V24=TSES24;
F1 = TSES;
/EQUATIONS
  V1  = 1F1 + E1;   V2  = *F1 + E2;
  V3  = *F1 + E3;   V4  = *F1 + E4;
  V5  = *F1 + E5;   V6 = *F1 + E6;
  V7 = *F1 + E7;    V8 = *F1 + E8;
  V9 = *F1 + E9;    V10 = *F1 + E10;
  V11 = *F1 + E11;  V12 = *F1 + E12;
  V13 = *F1 + E13;  V14 = *F1 + E14;
  V15 = *F1 + E15;  V16 = *F1 + E16;
  V17 = *F1 + E17;  V18 = *F1 + E18;
  V19 = *F1 + E19;  V20 = *F1 + E20;
  V21 = *F1 + E21;  V22 = *F1 + E22;
  V23 = *F1 + E23;  V24 = *F1 + E24;

/VARIANCES
E1 to E24 = *;
F1  = *;
/LMTEST
SET = PEE, GVF;
/Print
Fit = ALL;
Covariance = Yes;
/END
```

Appendix N
EQS Code for Model IV

```
/TITLE
  Model IV for Preservice Teachers (1-Factors Long-from 12 items
TSES)

/SPECIFICATIONS
  Data = 'D:\TSES Analysis\data\PreServiceShort.dat';
  VARIABLES=12 ; CASES=424;
  MATRIX=RAW;
  METHOD= ML;
  Analysis = COVARIANCE;
/LABELS
V1=t1; V2=t2; V3=t3; V4=t4; V5=t5;
V6=t6; V7=t7; V8=t8; V9=t9; V10=t10;
V11=t11; V12=t12;
F1 = TSES;
/EQUATIONS
  V1  = 1F1 + E1;
  V2  = *F1 + E2;
  V3  = *F1 + E3;
  V4  = *F1 + E4;
  V5  = *F1 + E5;
  V6  = *F1 + E6;
  V7  = *F1 + E7;
  V8  = *F1 + E8;
  V9  = *F1 + E9;
  V10 = *F1 + E10;
  V11 = *F1 + E11;
  V12 = *F1 + E12;
/VARIANCES
E1 to E12 = *;
F1  = *;
/LMTEST
SET = PEE, GVF;
/Print
Fit = ALL;
Covariance = Yes;
/END
```