

**MEDIATIONAL EFFECTS OF CLASSROOM MANAGEMENT ON TEACHER SELF-EFFICACY AND TEACHING EFFECTIVENESS**

A Dissertation Defense Presented to the  
Faculty of the College of Education  
University of Houston

In Partial Fulfillment  
of the Requirements for the Degree

Doctor of Philosophy

By

Lena Gerecht

August 2020

## Abstract

**Background:** Highly effective teachers have been shown to be the single most important school-based factor impacting student achievement. Despite this well-established relationship, it is still not clear why some teachers are better than others. Common and easy-to-measure characteristics like age, class size, post-baccalaureate coursework, professional degrees, professional exams, specialized certification, and participation in continued professional development have shown minimal to no significant impact on student achievement. Research on teacher's attitudes and beliefs in defining teacher quality has found mixed, but promising results for its relationship to self-efficacy. However, while self-efficacy and certain classroom practices have shown significant relations, self-efficacy has failed to yield significant correlations with teacher quality overall. Furthermore, few studies have explored potential mediating variables that impact the relationship between teacher self-efficacy (TSE) and teacher quality. One promising variable is classroom management, which has been cited as a foundational component of quality teaching and one that is typically skillfully practiced by teachers with high self-efficacy for teaching. **Purpose:** The present study investigated whether teachers' classroom management mediates the relationship between perceived teacher self-efficacy (TSE) and teachers' teaching effectiveness (TE). The study hypotheses are that (1) TSE will directly predict TE, (2) TSE will directly predict classroom management, and (3) TSE will have an indirect impact on TE via classroom management. Hypotheses were investigated with two different operationalizations of classroom management and TE (i.e., self and observer reports), resulting in four mediation models. **Methods:** Archival data was utilized from a three-year study that used mixed methods (e.g. self-report, archival records, ecological momentary assessment, and observational ratings) to assess teacher stress and its relationship to TE and student behavior. Participants were 202 6<sup>th</sup>-8<sup>th</sup> grade teachers from two urban school districts. Four multilevel path models were fit to examine if classroom management mediates the relationship between TSE predicting TE. Data are nested within teacher across time. **Results:** Partial support was found for all three hypotheses among the four models. TSE directly predicted both self-reported TE and classroom management. TSE did not

predict observer-reported TE and classroom management. Regarding the third hypothesis, the two models based on self-report provided support for the indirect relationship of TSE on TE when classroom management was self-reported. Additionally, in each of the four model's, classroom management, whether self- or observer-reported always significantly predicted self- and observer ratings of TE.

## Table of Contents

Chapter	Page
I. Introduction.....	1
II. Literature Review.....	13
a. Teacher Quality.....	13
b. Measuring Teaching Effectiveness.....	25
c. Classroom Management.....	31
d. Teacher Self-Efficacy.....	33
e. Gaps in the Literature & Aims of the Proposed Study.....	46
III. Methodology.....	49
a. Participants.....	49
b. Study Design.....	49
c. Measures.....	50
d. Procedures.....	53
e. Data Analysis Plan.....	53
IV. Results.....	55
a. Descriptive Statistics.....	55
b. Hypotheses Results.....	72
V. Discussion.....	74
a. Strengths and Limitations.....	77
b. Implications.....	78
c. Future Directions.....	79
References.....	81
Appendix A.....	107

## **Chapter I:**

### **Introduction**

Since the publishing of *A Nation at Risk* by the National Commission on Excellence in Education (NCEE) in 1983, state education policy makers and reformers have suggested many different ideas and initiatives to make our nation's schools more effective and internationally competitive. The federal government responded in 2001 with the passage of No Child Left Behind (NCLB, 2001) that directed this reform movement towards a focus on accountability for student achievement. This meant that districts were required to report disaggregated annual test scores to become a part of public record from each of their schools. These annual reports had serious implications for districts. The anticipated benefit of this requirement was that assessment results would directly translate into instructional accommodations. Along with a keener focus on academic standards, standards-based instruction, and data-driven instructional techniques, the legislation called for more teacher accountability and for a push for highly qualified teachers in every classroom across the nation. By addressing these issues and meeting these demands states and districts would be steadily increasing student achievement goals.

With the increase in pressure from the federal government, reformers and state policy makers began to place greater attention on the significance of teacher quality as a critical means of promoting student achievement. However, despite research consistently showing that the quality of a teacher is the single most important school-based factor in student achievement, determining characteristics that define quality have remained elusive (Hanushek & Rivkin, 2010; McCaffrey, Lockwood, Koretz, & Hamilton, 2003). As such, policy makers, researchers, and school administrators have grappled with how to define, measure, and evaluate teacher effectiveness.

Despite lacking a good measure of teacher quality, and because of the increasing pressure from the federal government for more accountability, there has been no shortage of standards to outline descriptions of accomplished teaching practices, certification systems to qualify teachers, and evaluation systems to assess teacher effectiveness in the classroom. Furthermore, since these processes most typically are the responsibility of each state, there is considerable variation. Thus, it is unsurprising that the success and utility of certification and evaluative measures of teacher effectiveness has been mixed (Clare, Valdés, Pascal, & Steinberg, 2001; Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; Darling-Hammond, Holtzman, Gatlin, & Heilig, 2005; Goe, Bell, & Little, 2008; Goldhaber & Brewer, 2000; Harris, Ingle, & Rutledge, 2014; Kane, Rockoff, & Staiger, 2008; Kyriakides, Campbell, & Christofidou, 2002; Little, Goe, & Bell, 2009; Milanowski, Kimball, & White, 2004; Qu & Becker, 2003; Tucker, Stronge, Gareis, & Beers, 2003).

A better understanding of what constitutes teacher quality has significant implications for decision making regarding the preparation, recruitment, compensation, inservice professional development, and evaluation of teachers (Covino & Iwanicki, 1996; Dunn & Mulvenon, 2009; Goldhaber & Brewer, 2000; Milanowski et al., 2004; Stronge, Ward, & Grant, 2011; Tucker et al., 2003). Hence, it becomes increasingly important that researchers clarify established variables as well as explore alternative ones that could potentially reveal relations to teacher quality. The body of research on teacher quality, while mixed, has revealed some promising relations to certain characteristics and instructional practices that warrant further investigation. These include self-efficacy, a motivational construct, and both classroom management and organization.

## **Teacher Quality: A Brief Overview**

Researchers have determined that producing gains in student achievement relies heavily on improving the classroom teacher (Darling-Hammond, 2003; Hanushek, Kain, O'Brien, & Rivkin, 2005; Hanushek & Rivkin, 2010; Nye, Konstantopoulos, & Hedges, 2004; Rice, 2003). The value of having a quality teacher, as measured by their ability to improve student achievement, is greater than any other school input and can exceed one grade-level of achievement per year (Darling-Hammond, 2003; Goldhaber & Anthony, 2003; Hanushek & Rivkin, 2010). For example, students starting at the same baseline achievement at the beginning of the school year can make significant gains in achievement at the end of a single academic year due solely to the quality of the teacher to which they were assigned. As such, teacher quality is the focus of many reforms and policies designed to improve student achievement in schools. However, when defining what differentiates low quality teachers from high quality teachers, many states and districts tend to examine easy-to-measure teacher characteristics like post-baccalaureate coursework, subject-matter education, professional degrees, professional exams, years of experience, specialized certification, and evidence of participation in continued learning and professional development (Goe & Stickler, 2008; Haney, 1987; Hanushek et al., 2005). While these are important considerations upon entry into the profession, only two—content area expertise in mathematics and years of experience during a teacher's first five years—have been shown to have any direct or significant impact on student achievement (Berry, 2010; Cohen, Raudenbush, & Ball, 2003; Darling-Hammond, 2003; Darling-Hammond & Sykes, 2003; Hanushek, 2011; Hanushek & Rivkin, 2010; Wenglinsky, 2000).

Policies such as those created by the NCLB resulted in teacher assessments that have relied heavily on standardized test scores of proficiency or growth indicators to measure teacher

effects on student achievement (Goe & Stickler, 2008; Hanushek, 2011; Hanushek & Rivkin, 2010; Rivkin, Hanushek, & Kain, 2005). These measures, also known as value-added measures, involve multilevel modeling techniques that are used to infer quality from residual estimates of student test scores after controlling for student demographic characteristics (Goe & Stickler, 2008; McCaffrey et al., 2003; Newton, Darling-Hammond, Haertel, & Thomas, 2010). As is the case with teacher quality, studies that examine the use of value-added measures consistently indicate that the majority of variation in teacher effectiveness is due to unobserved variables. This highlights a critical problem with using methods that depend solely on standardized tests to measure teacher effectiveness. These tests, although measuring student acquisition of specific content standards, are not developed to discern variations in instruction or to differentiate teacher contributions to student learning (Goe & Stickler, 2008; Hanushek & Rivkin, 2010; McCaffrey et al., 2003; Newton et al., 2010). Thus, while student achievement scores help to identify these teachers, they do not reveal how or why these teachers are effective at improving student outcomes (Stronge et al., 2011).

Replacing NCLB in 2015, the Every Student Succeeds Act (ESSA) allowed states more freedom to evaluate student growth and teacher quality. Subsequent to passage of ESSA, schools have the option to rely less heavily on annual student test scores, and to incorporate a variety of measures (Black, 2017; Mathis & Trujillo, 2016; Podolsky, Kini, Bishop, & Darling-Hammond, 2017). Reduced reliance on test scores was in response to wide skepticism of both teachers and policymakers that teachers working in the most demanding instructional environments will be penalized because of factors affecting student learning that are not under the teacher's direct control (Jennings & Rentner, 2006). Furthermore, studies of value-added modeling as a measure of teacher effectiveness, such as a study by Newton et al. (2010), found



that using test scores alone can both overestimate and underestimate a teacher's effectiveness based on the particular course, grade, or students the teacher is assigned. Thus, states, while still relying primarily on annual assessments, in addition can incorporate a variety of measures such as classroom observations when evaluating student achievement and teacher quality. Even prior to ESSA, most districts utilized other methods of assessment including classroom observation and some self-report measures (Goe et al., 2008). Classroom observations provide information on the practices or teacher behaviors that could potentially increase student learning and, with few exceptions, have been found to significantly predict student achievement (Rockoff & Speroni, 2010).

The classroom processes and instructional practices that teachers perform to impact student learning is also known as a teacher's teaching effectiveness (TE). Structured observation can elucidate differences in TE which has strong implications for training, evaluation, and the continuing education of teachers. Further, teacher self-report measures can be used to assess TE by self-evaluations of what they are doing in the classroom. However, in general, these self-report measures have presented with mixed results in the literature as to their reliability and validity to TE and teacher quality in general (Goe et al., 2008; Hsu, 2005; Keidel, 2014). Moreover, these measures are often retrospective and are thus subject to retrieval and appraisal biases (Kihlstrom, Eich, Sandbrand, & Tobias, 2000; Shiffman, Stone, & Hufford, 2008).

Ecological Momentary Assessment (EMA), or repeated collection of real-time data on momentary states within the environment, is a promising alternative methodology to evaluate TE. EMA data collection allows for immediate reporting that minimizes retrieval biases. Repeated EMA assessments reveal within-person variability over time. The collection of contextualized data increases the ecological validity and provides valuable information on

situation-specific malleable factors for targeted interventions. With the succession of ESSA, it is likely that there will be increased efforts to utilize multiple methods to assess student growth and TE. In addition, these alternative methods can provide invaluable information that can be utilized by districts to train, identify, and improve teacher quality and thus, improve student outcomes.

### **Classroom Management and Organization**

Historically, research on characteristics of quality teachers has been mixed and inconclusive; however, observed teaching practices within the classroom reveal a clearer picture of what makes a qualified teacher. In particular, when teachers are reported to be skilled practitioners of classroom management and organization, they generally tend to provide high quality instruction (Goldhaber & Anthony, 2003; Stage & Quiroz, 1997; Stronge, 2007).

Classroom management and organization, or the way teachers manage the physical and behavioral aspects of the classroom environment, is one way to promote more class time spent working productively on academic tasks, which, in turn, is associated with higher academic achievement (Marzano, Marzano, & Pickering, 2003; Wesley & Vocke, 1992). Stronge et al. (2011) found that the top quartile of teachers, as measured by their student achievement, scored significantly higher in establishing routines, using time efficiently and effectively, and monitoring student behavior—all practices of good classroom management. Studies on teacher quality found that those who were rated to be high quality teachers, in general, were found to be better managers and organizers, encouraged students to have more responsibility in the classroom, were consistently well-prepared, had predictable daily classroom routines, were able to transition between activities quickly, handled routine tasks promptly and efficiently, and prevented or minimized interruptions (Stronge et al., 2011; Wang, Haertel, & Walberg, 1993; Wharton-McDonald, Pressley, & Hampston, 1998). Thus, the importance of classroom

management and organization for quality of teaching cannot be overstated; if effective classroom management is not present, it is often a strong indicator that the teacher is unsuccessful in the classroom (Freiberg, Huzinec, & Templeton, 2009; Keidel, 2014).

Classroom management is rated as one of the most challenging aspects of teaching for novice and veteran teachers alike (Carr, 2013; Dunn & Mulvenon, 2009; Hicks, 2012; Stronge, 2006). Additionally, many teachers feel underprepared to take on the challenges of establishing procedures and managing student disciplinary concerns. While troubling, these findings are not surprising given that many teachers enter the field receiving minimal to no training in classroom management techniques (Lavay, Henderson, French, & Guthrie, 2012; Oliver & Reschly, 2010; Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008; Stough, 2006; Wesley & Vocke, 1992).

Research has shown that teachers who report having greater self-efficacy tend to have positive classroom management experiences and student interactions (Klassen, Tze, Betts, & Gordon, 2011). Conversely, teachers who perceive themselves as being less efficacious in classroom management are more likely to experience negative emotions (i.e. depression and anxiety) and burnout. Given these findings, classroom management also has been found to be a strong contributing factor of teacher attrition (Dicke et al., 2014; Goldring, Taie, & Riddles, 2014).

### **Self-Efficacy**

Perceived self-efficacy describes a person's self-conception of their own performance capability in one or more distinct areas of functioning (Bandura, 1997). Rooted in Bandura's (1977) social cognitive theory, perceived self-efficacy is a central determinant of individuals' motivation, with particular importance for decisions to engage and persevere in a specific task or action. One of the basic principles of self-efficacy theory is that people are more likely to

engage in activities in which they have high self-efficacy and less likely to engage in those they do not. Performance and motivation are in part determined by how effective people believe they can be. Thus, highly efficacious people tend to be more effortful, invest more time, set higher goals, persist longer in the face of obstacles, and are more resilient when faced with setbacks.

Self-efficacy as it relates to the task of teaching is known as teacher self-efficacy (TSE) or teachers' beliefs in their ability to organize, plan, and carry out activities required to influence students' valued educational goals and outcomes (Klassen et al., 2011; Ross, 1998). TSE has been related to teacher behavior in the classroom. TSE influences more than students' behavior. TSE is correlated with self-efficacy beliefs which are associated with student motivation and achievement (Ross, 1992; Tschannen-Moran, Hoy, & Hoy, 1998). Although modeling is an important consideration in self-efficacy theory (Bandura, 1997; Hagen, Gutkin, Wilson, & Oats, 1998; Rubeck & Enochs, 1991), a growing body of research has supported the notion that TSE beliefs influence student outcomes via instructional quality which is decisive for students' achievement outcomes (Emmer & Stough, 2001). The majority of studies on TSE cite that TSE is related to specific teachers' behaviors as well as the practices and strategies they use in the classroom (Abu-Tineh, Khasawneh, & Khalaileh, 2011; Cason, 2018; Dicke et al., 2014). However, results have been mixed regarding TSE's relationship to overall levels of teacher quality as well as student achievement. In a review of the literature, Holzberger, Philipp, and Kunter (2013) state that this is likely in part due to the significant limitations of previous research. These limitations include lack of longitudinal studies, the fact that teacher quality was almost exclusively measured using self-reports rather than other data sources such as student ratings or classroom observation. Exclusive reliance on self-report introduces method bias, generally in the direction of an overly positive self-appraisal. Furthermore, studies of TSE tend

to be small, with limited power to find effects. Also, very few studies to date have explored potential mediating variables that may impact the relationship between TSE and teacher quality.

### **Problem Statement**

There is considerable variability in student outcomes that is empirically connected with specific teachers. Some teachers achieve half of the growth expected in a school year, whereas others get their students to grow at rates that are 1.5 or even twice the school average (Darling-Hammond, 2000). Despite this ubiquitous and well established relationship between teacher quality and student success, it is not clear why some teachers are better than others (McCaffrey et al., 2003; Partee, 2012). The body of research on teacher quality since the NCLB act has largely focused on student achievement gains at the expense of other measures such as observation or self-report that could elucidate the attributes (e.g. characteristics and qualifications) and processes or a teacher's TE (e.g. planning, classroom management, use of instructional strategies, interactions with students) of high quality teachers. Research on teacher's attitudes and beliefs in defining teacher quality has found mixed, but promising results for its relationship to self-efficacy. However, while self-efficacy and certain classroom practices such as classroom management have shown significant relations, self-efficacy has failed to yield significant correlations with teacher quality overall (i.e., the impacts on student achievement) (Caprara, Barbaranelli, Steca, & Malone, 2006; Chong, Klassen, Huan, Wong, & Kates, 2010; Hardré et al., 2006; Tschannen-Moran & Hoy, 2007).

Failing to find a significant relationship between TSE and student achievement is likely due, at least in part, to the significant limitations of previous research including a lack of longitudinal studies, small sample sizes, and differences in the conceptualization and measurement of teacher quality. The majority of studies on TSE and teacher quality rely almost

exclusively on retrospective self-report measures of teacher quality. Reliance solely on self-report rather than other measures of teacher quality (i.e. observations or EMA) is problematic in that it introduces method biases. Furthermore, very few studies to-date have explored potential mediating variables that impact the relationship between TSE and teacher quality. One promising variable is classroom management, which has been cited as a foundational component of quality teaching and one that is typically skillfully practiced by teachers with high self-efficacy for teaching (Brophy, 1988; Holzberger et al., 2013).

### **Purpose of the Study**

The purpose of this study is to understand if classroom management mediates the relationship between TSE and a significant and measurable component of overall teacher quality, TE. Additionally, this study seeks to determine whether or not the mediation model generalizes across different raters or is rater dependent (see Figure 1). This study could provide school districts, policy makers, and educators with valuable information on characteristics and processes that impact teacher quality. Furthermore, this study could provide guidance to school staff and administrators in teacher preparation and in remediating teachers' TE. Another benefit is that this study may add to the existing body of research on TSE by further clarifying the mixed evidence of relations between TSE and teacher quality. Finally, depending on the outcomes of this study, whether or not a mediation model is applicable, this may open a promising avenue of TSE research wherein the specific domains of TSE can be analyzed to determine their relations within this model. This has important implications for teacher training, as developing TSE for classroom management could become a critical goal in teacher preparation and professional development.

### **Hypotheses**

The hypotheses examined in this study are as follows:

1. TSE will directly predict teaching effectiveness.
  - a. TSE will directly predict self-reported teaching effectiveness.
  - b. TSE will directly predict observer-reported teaching effectiveness.
2. TSE will directly predict classroom management.
  - a. TSE will directly predict self-reported classroom management.
  - b. TSE will directly predict observer-reported classroom management.
3. TSE will have an indirect impact on teaching effectiveness via classroom management.
  - a. TSE will have an indirect impact on self-reported teaching effectiveness via self-reported classroom management.
  - b. TSE will have an indirect impact on self-reported teaching effectiveness via observer-reported classroom management.
  - c. TSE will have an indirect impact on observer-reported teaching effectiveness via self-reported classroom management.
  - d. TSE will have an indirect impact on observer-reported teaching effectiveness via observer-reported classroom management.

These questions will be addressed utilizing archival data from a three year study of the school years 2011-2012 and 2012-2013. This study used multiple methods (e.g., self-report, archival records, ecological momentary assessment (EMA), and observational ratings) to assess teacher stress and its relationship to teacher effectiveness and student behavior. Participants were 202 sixth through eighth grade teachers from 22 middle schools in two southern urban school districts in the United States. These archival data provide longitudinal data of class period

specific self and observer ratings of teacher's classroom management, organizational quality, and overall instructional quality. Thus, four multilevel path models will be fit to examine if classroom management mediates the relation between TSE and teaching effectiveness.



## Literature Review

### Teacher Quality

Schools make a difference in student learning and a substantial portion of that difference is directly attributable to teachers (Darling-Hammond, 2000, 2003; Hanushek et al., 2005; McCaffrey et al., 2003; Rowan, Correnti, & Miller, 2002; Stronge et al., 2011). Teacher quality is a term used in research and educational policy whereby teacher related attributes are evaluated to help determine how well teachers influence student learning. Teacher quality assessments attempt to peer into the “black box” of teaching and determine why some teachers are more effective than others. A conundrum in the current literature is that there are well-established teacher effects, such that student achievement varies with teachers, but no general consensus on how to measure teacher quality.

Assessing teacher quality is a complex issue fraught with methodological and terminological uncertainty. In the literature, teacher quality typically has been conflated as a term that means what teachers possess in the way of certification or training, what teachers practice in the classroom, or that one assumes the other (Ahn & Choi, 2004; Darling-Hammond, 2000; Goe, 2007; Goldhaber & Anthony, 2003; Hanushek, 2011; Kennedy, 2010). Furthermore, teacher quality also has been represented as student learning gains on standardized tests (e.g., student outcomes), teacher qualifications (e.g., teacher education, characteristics, and experience), and the quality of provision of instruction within the classroom (e.g., TE). As a means of making sense of the ways in which teacher quality has been represented in the literature, Goe (2007) created a framework for teaching quality that consists of three categories that represent ways of evaluating teacher quality. These include inputs, processes, and outputs.

**Inputs.** Inputs include both teacher characteristics and teacher qualifications. Teacher qualifications refer to preexisting resources that are applied to the classroom context (Goe, 2007). These preexisting resources are also considered important in establishing who is and who is not qualified to teach and thus are used as a reference in order to make hiring decisions. These can include coursework, subject matter education, degrees, grades, test scores, certification, credentials, and experience. A great deal of research has been dedicated to examining the relationship between teacher qualifications and student learning gains. However, findings have been inconsistent at best (Betts, Zau, & Rice, 2003; Darling-Hammond et al., 2005; Darling-Hammond & Sykes, 2003; Goldhaber & Brewer, 2000; Hanushek et al., 2005; Nye et al., 2004; Rice, 2003; Rivkin et al., 2005). For example, teacher certification has generally found little support in the literature for affecting student educational goals and often fails to differentiate between high and low quality teachers (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2006; Glazerman, Mayer, & Decker, 2006; Guskey & Sparks, 2004; Hanushek et al., 2005; Hattie, 2008; Kane et al., 2008; Qu & Becker, 2003).

In a meta-analysis of teacher inputs by Qu and Becker (2003) the authors examined the results from 24 studies comparing certified teachers to teachers who had obtained other kinds of certificates (e.g. alternative certification and emergency certification). The authors found a negligible effect size of a teachers' pathway to certification on student outcomes ( $d = 0.10$ ). In an experimental study by Glazerman et al. (2006), students were randomly assigned to 44 teachers with emergency licenses and 56 trained (certified) teachers. They found no differences in reading ( $d = .03$ ) and small differences in mathematics ( $d = .15$ ). Thus, training inputs appear to be a small contributor to quality assessments.

Subject matter knowledge is another variable of teacher qualification that is plausibly related to teacher quality. In general, studies have found mixed results for subject matter knowledge. Researchers have reported positive and negative, small, and statistically nonsignificant relations between teachers' knowledge of their subject matter they teach with student achievement gains (Ahn & Choi, 2004; Andrews, 1980; Darling-Hammond et al., 2005; Haney, 1987; Hanushek et al., 2005; Hattie, 2008; Rivkin et al., 2005).

In a meta-analysis of subject matter by Ahn and Choi (2004), the authors examined 27 studies relating teachers' subject matter knowledge to student mathematics achievement. The authors found a very small ( $r = .06$ ) mean correlation. Furthermore, these effects were small for both elementary ( $r = .05$ ) and high school ( $r = .11$ ) teachers. Furthermore, studies of teachers' scores on subject matter tests of the National Teacher Examinations have also found no consistent relations between these scores and teacher quality as measured by student outcomes or supervisor ratings (Andrews, 1980). Thus, teachers' expertise in specific teaching domains does not seem to be robust predictor of teaching quality.

Among all teacher qualifications, only the amount of experience has consistently been related to gains in student achievement (Covino & Iwanicki, 1996; Darling-Hammond et al., 2005; Dee, 2004; Hanushek et al., 2005). However, this relationship becomes statistically nonsignificant after the first five years of experience as the studies show that a teacher with five years of experience performs similarly to teachers with 25 years of experience with respect to student learning gains (Darling-Hammond et al., 2005; Hattie, 2008; Kane et al., 2008; Staiger & Rockoff, 2010).

**Teacher characteristics.** The second category of inputs in the Goe (2007) model of teacher quality is teacher characteristics. Teacher characteristics include attitudes and attributes

(i.e. expectations of students, locus of control, perceived self-efficacy), as well as immutable characteristics such as gender and race. Many of the findings of assigned characteristics such as gender and ethnicity have been mixed (Dee, 2004, 2007; Egalite, Kisida, & Winters, 2015; Saft & Pianta, 2001). However, several studies have found positive relations between ethnicity and student achievement (Clotfelter, Ladd, & Vigdor, 2010; Dee, 2004, 2007; Hanushek et al., 2005). For example, Hanushek et al. (2005), using matched panel data from a large school district in Texas, found an inconsistent direct relation between student gains and race. Further, when students and teacher were matched by ethnicity there were positive effects on student achievement gains.

In a study using random assignment of students, Dee (2004) found that African American students taught by same race teachers for a year had three to six percentile increases in both math and reading on the Stanford Achievement Test. Furthermore, the author also found that Caucasian students taught by Caucasian teachers was associated with a four to five percentile point increase in math scores. However, other studies have found no relations between these variables (Clotfelter et al., 2010; Darling-Hammond, 2000; Ehrenberg, Goldhaber, & Brewer, 1995). For example, Ehrenberg et al. (1995), utilizing the National Education Longitudinal Study of 1988, sought to determine whether teacher race, gender, and ethnicity mattered in terms of student learning. The authors found little evidence of meaningful associations between student achievement and these teacher characteristics. Overall, their related effects were small and inconsistent across grade-levels and subjects taught.

To investigate teacher quality, researchers have also examined teachers' attitudes and beliefs as they relate to the task of teaching. Measures of teachers attitudes and beliefs in the literature typically include measures of motivation, self-efficacy, and expectations for students

(Goddard, Hoy, & Hoy, 2000; Guo, Piasta, Justice, & Kaderavek, 2010; Palardy & Rumberger, 2008; Van den Bergh, Denessen, Hornstra, Voeten, & Holland, 2010). Regarding teachers attitudes, Palardy and Rumberger (2008), found both reading and math achievement gains tended to be lower among teachers with low expectations for their students. Similarly, Van den Bergh et al. (2010) found a positive relation between teachers' negative expectations and math test scores of ethnic minority students.

Teacher Self-Efficacy (TSE) and collective teacher self-efficacy have also been found to be related to aspects of instructional quality and achievement gains (Goddard et al., 2000; Guo et al., 2010; Woolfolk, Rosoff, & Hoy, 1990). Studies on TSE and student achievement generally suggest that TSE likely predicts students' achievement through classroom practices in elementary, middle, and high school (Caprara et al., 2006; Chang, Lin, & Song, 2011; Chong et al., 2010; Hardré et al., 2006; Jimmieson, Hannam, & Yeo, 2010; Mohamadi & Asadzadeh, 2012; Mojavezi & Tamiz, 2012; Woolfolk Hoy, Hoy, & Kurz, 2008). However, the coefficients in many of these studies were generally modest ranging from .02 to .78 with a median of .27.

Collective teacher self-efficacy is another construct thought to be related to student learning gains that has gained increasing attention in the past two decades. Collective teacher self-efficacy, or the teaching faculty's beliefs in its collective ability to carry out teaching tasks to promote student achievement, relates to how likely a collective is to put forth sustained effort and persistence towards a common goal (Goddard et al., 2000). In a study by Goddard et al. (2000) the authors, using hierarchical linear modeling found significant associations between student achievement and collective teacher self-efficacy. Specifically, the authors found that collective teacher self-efficacy explained more than half of the between school variance in mathematics (53.27%) and reading (69.64%). Taken together, the study of TSE, both

individually and collectively, is one of the more promising areas of investigation in the search for characteristics that are able to measure and predict teacher quality.

**Processes.** Teacher processes include teacher practices both in and outside the classroom (Goe, 2007). These include planning, classroom management, use of instructional strategies, presentation of instruction, and teachers' interactions with students. This is often referred to as "instructional quality" or "teaching effectiveness (TE)," an active and integral component of teacher quality (Guskey, 1988). Studies of TE employ observational protocols, interviews, work examples, and self-report measures to evaluate teacher behaviors and practices within the classroom (Goe, 2007; Pianta & Hamre, 2009; Stronge, 2006). Several common standards-based teacher evaluation systems have been used in the research and in schools to examine the quality of teachers' provision of instruction (Goe et al., 2008; Pianta & Hamre, 2009). In general, these metrics tend to vary in the literature, however, most evaluate common foundations of instruction that include planning, classroom management, use of instructional strategies, presentation of instruction, and teachers' interactions with students.

Many elements of the teaching process have been linked to TE including keeping students focused and engaged, asking questions, and the content of the curriculum (Covino & Iwanicki, 1996; Hattie, 2008; Klassen et al., 2011; Pianta, Hamre, & Allen, 2012; Stronge, 2007; Waxman, Hilberg, & Tharp, 2004). However, there is no single classroom practice that is necessarily effective with all subjects, grade levels, and students, as such, teachers must be both flexible and opportunistic.

Studies on teachers who utilize a variety of teaching strategies that are appropriate to the instructional goals of the class, are more likely to produce student learning gains (Stronge, 2007; Stronge et al., 2011). By successfully employing a variety of strategies they are able to reach

more students by addressing a wider range of student interests and learning needs (Schumacher, Grigsby, & Vesey, 2015; Tomlinson, 2000). Furthermore, effective teachers provide appropriate strategies based on time, resource constraints, and content and are able to adjust instruction appropriately. These instructional strategies can include hands-on learning activities, integrating inquiry-based problems, critical thinking skills, mastery learning, and testing in daily lessons (Ku, Ho, Hau, & Lai, 2014; Wijnia, Loyens, van Gog, Derous, & Schmidt, 2014).

A teacher's TE has been identified as one of the most significant variables that effects student cognitive, behavioral, and affective outcomes (Kunter et al., 2013; Palardy & Rumberger, 2008; Porter & Brophy, 1988; Reyes, Brackett, Rivers, White, & Salovey, 2012; Stronge, 2007; Stronge et al., 2011). In a meta-analyses by Hattie (2008), the author synthesized behaviors and practices of teachers that were related to student achievement gains including cooperative learning, mastery learning, challenging learning goals, direct instruction, feedback, problem-solving instruction, and teacher clarity. These effect sizes ranged from .41 to .75 with teacher clarity being the most significant strategy predicting student outcomes. Additionally, coaching and modeling have been shown to be effective practices that teachers employ in the classroom (Dukerich, 2015; Park & Logsdon, 2015). Furthermore, these teachers know how to differentiate instruction based on individual student needs (Cawelti, 2004; Tomlinson, 2000; Tomlinson & Moon, 2013).

Students differ in their levels of knowledge and understanding, thus the teacher's provision of differentiated instruction is essential for them to access a variety of learners. This is done by presenting information in a way that is challenging yet appropriate to the student, which involves adapting assignments, and teaching organizational skills (Artelt & Schneider, 2015; Conner, Singletary, Smith, Wagner, & Francisco, 2014; Cotton, 2002; Covino & Iwanicki,

1996). An additional way for the curriculum to be accessible to a variety of learners at different levels is via grouping strategies. Grouping strategies have also been shown to be a common practice among effective teachers; grouping strategies can include flexible grouping, ability grouping and cooperative grouping (Hoffman, 2002; Melsner, 1999; Vaughn, Bos, & Schumm, 1997).

High quality teachers also place greater emphasis on their students understanding and demonstrating the meaning of instructional content rather than facts alone (Demmon-Berger, 1986; Donovan, Bransford, & Pellegrino, 1999; Good & Lavigne, 2017; Orthner et al., 2010). Effective teachers stress the significance of higher mental processes such as analytical thinking skills, creativity, and problem-solving techniques. Teachers facilitate deeper learning through use of high-quality questioning that allows students the opportunity to think critically (Brophy, Good, & Wittrock, 1986; Cawelti, 2004; Marshall & Smart, 2013; Weiland, Hudson, & Amador, 2014).

Student engagement is a prerequisite to student learning and thus essential to academic achievement (Christenson, Reschly, & Wylie, 2012; Reyes et al., 2012). Engaged students participate, exert effort, and are motivated to learn whereas students with low engagement are more likely to be disruptive, have lower grades, and are less likely to aspire to high academic goals (Fredricks, Blumenfeld, & Paris, 2004; Kaplan, Peck, & Kaplan, 1997; Reyes et al., 2012; Stronge, 2007). To address students' off-task behaviors, high quality teachers will modify instruction to make it more engaging and employ a variety of organizational and managerial strategies that promote on-task behavior. However, the most important means of promoting student engagement is through the classroom climate. Studies have shown that teachers who create a caring, supportive, and organized environment in which expectations are high, clear, and



fair are more likely to have engaged students (Fredricks et al., 2004; Kaplan et al., 1997; Klem & Connell, 2004). Furthermore, positive classroom climates are related to reduced emotional problems and enhanced academic achievement, whereas negative classroom climates are related to increased bullying, social-emotional maladjustment, and aggression (Djigic & Stojiljkovic, 2011; Evans, Harvey, Buckley, & Yan, 2009). In a meta-analysis of 119 studies, Cornelius-White (2007) examined the impact of learner-centered teacher-student relationships on student educational outcomes. Those students who identified their relationships with their teachers as warm, empathic, genuine, non-directive, and encouraging of learning had significant relations to positive cognitive and behavioral outcomes.

Along with developing positive relationships with their students, teachers with high TE create a positive and supportive learning environment via classroom management and organization (Fredricks et al., 2004; Kaplan et al., 1997; Reyes et al., 2012; Stronge, 2007). Teachers are able to maintain control over their classroom by establishing routines and procedures that limit the time taken away from class and student behavioral concerns. Having a high quality, well managed classroom assists teachers in maintaining variety and momentum in their instructional delivery, monitor and respond to student activity, and in turn, prevent and address discipline problems (Beaty-O’Ferrall, Green, & Hanna, 2010; Emmer & Sabornie, 2013; Stronge, 2007). Having effective classroom management and organization skills means fewer disruptions, less time spent discipline tasks, and fewer issues during unstructured and transition time (Conner et al., 2014; Emmer & Stough, 2001).

Many studies have shown the importance of classroom management as being an essential part of TE (Corbett & Wilson, 2002; Kane, Taylor, Tyler, & Wooten, 2011; Stronge et al., 2011; Wang et al., 1993). Teacher organization of the learning environment is equally important in

creating and sustaining learning within the classrooms. This involves the teacher physically orienting the classroom for instruction, framing lessons in a coherent and logical manner, and preparing and organizing materials (Cornell & Mayer, 2010; Stronge, 2007). Characteristics of an organized classroom include established routines and procedures, having materials prepared and ready, arranging classroom furniture to define learning areas, and communicating to students the organization of space where materials are to be stored (Bain & Jacobs, 1990; Guardino & Fullerton, 2010). Thus, the many teaching practices that promote positive student engagement and good behavior seem to be critical to quality teaching.

**Outputs.** The final dimension of Goe's framework pertains to outcomes of teacher quality as measured by teachers' impacts on student learning gains (Goe, 2007). Typically, these outputs are determined by performance evaluations and growth estimates. Performance evaluations are conducted by either external or internal evaluators (e.g. school principals) and can vary widely in the domains they evaluate. Generally school evaluations consider multiple elements including classroom observations, lesson plans, work examples, and other documents (Goe et al., 2008; Pianta & Hamre, 2009). Historically variables to evaluate teacher quality have differed significantly by district and state, making the existing research difficult to compare or aggregate (Medley & Coker, 1987; Tucker et al., 2003). Furthermore, studies of teacher evaluations conducted by school administrators have found mixed evidence for its relationship to student achievement (Jacob & Lefgren, 2008; Medley & Coker, 1987; Milanowski et al., 2004; Murnane, 1975). In many of these evaluations, the quality of instruction is not directly measured because too few assessment tools exist that have the potential to measure the quality of classroom practice on a large-scale basis. Most are time consuming and expensive to conduct.

Given the abundance of student standardized test score data and increased accountability standards, researchers have turned to advanced statistical techniques to determine the effects of teacher quality on student learning (Grossman, Loeb, Cohen, & Wyckoff, 2013). One common objective of measuring a teacher's contribution to student achievement is to isolate an estimate of a teacher's contribution to student achievement separate from the student characteristics, class, school, and other contributors (Hanushek & Rivkin, 2010; Kimball, White, Milanowski, & Borman, 2004; Milanowski et al., 2004). One way to do this is through modeling growth in achievement, as opposed to achievement levels also known as value-added measures (Grossman et al., 2013).

Many states have legislation requiring the use of value-added models to comprise the majority of teacher evaluations, which are the basis for high-stakes decisions about promotion, tenure, dismissal, and compensation (Harris et al., 2014). Value-added measures aim to determine the extent that students' academic progress from one year to the next is attributable to his or her teacher, rather than factors outside the teachers control (e.g. community, school history, family, socio-economic status, etc.) (Jacob & Lefgren, 2008; Popham, 2006). This is accomplished by (1) establishing a baseline of previous performance on an assessment, (2) collecting data on student characteristics and school-wide factors (uncontrolled), and (3) using prior performance, current school year performance, student characteristics, and classroom and school-wide factors for all students. This information is entered into a statistical model that predicts the performance that a typical student with similar characteristics would be expected to have on an end-of year assessment. Finally, this data is then compared to the student's actual performance to the model's prediction for a typical student with these characteristics. The process is repeated for each student assigned to a given teacher and the results of these

comparisons are averaged and generate a value-added score, or teacher effect, for each teacher. If a teacher's students consistently outperform the model's predictions, this teacher is assigned a higher value-added score. If the teacher's students underperform relative to the prediction, the teacher is assigned a lower value-added score.

The benefit of using value added measures is that by accounting for prior test scores, value-added measures help account for the nonrandom assignment of students to teachers and yield a less biased measure of teacher effectiveness (Guarino, Reckase, & Wooldridge, 2010; Harris & Sass, 2006; Kane & Staiger, 2008; Todd & Wolpin, 2003). In general, studies using value-added measures have demonstrated that there are significant differences in teacher effectiveness (Hanushek & Rivkin, 2010; Milanowski et al., 2004). For example, using a matched panel data from UTD Texas Schools Project, Rivkin et al. (2005), sought to examine the impact that teachers have on student reading and mathematics achievement. They found that, when using an empirical estimation of within-school variance of teacher effectiveness, a one standard deviation increase in average teacher quality raised the average achievement in that grade by at least .11 standard deviations in mathematics and .10 standard deviations in reading.

Despite some support for these models, there are several limitations to using value-added measures to determine teacher quality. Value-added measures assume that student learning is measured well by given standardized tests, which is not always the case. Value-added measures presumed that growth in scores is influenced by the teacher alone, and is independent from the growth of classmates and other aspects of the classroom context (Darling-Hammong, Amrein-Beardsley, Haertel, & Rothstein, 2012). Furthermore, value-added measures provide no information as to how and why some teachers are more effective than others (Hanushek &

Rivkin, 2010). Thus, value-added measures by themselves are an inadequate replacement for traditional teacher evaluation (Hanushek & Rivkin, 2010).

### **Measuring Teacher Effectiveness**

TE is malleable factor that has a large influence on students' achievement (Hattie, 2009). In order to improve the quality of instruction it is important to identify the strengths and weaknesses of teachers concerning their actual teaching practices or TE (Rivkin et al., 2005). According to Waxman et al. (2004), there are four key foci when investigating TE: (1) description of instructional practices, (2) improving teacher's classroom instruction, (3) investigating instructional inequities for different groups of students, and (4) evaluating and improving teacher education programs. Evaluations of teachers' classrooms by school principals and external reviewers are the most common metrics along with teacher and student surveys (Darling-Hammond et al., 2012; Harris et al., 2014). Currently administrators utilize multiple sources of data that are collected at several points throughout the year (Darling-Hammond et al., 2012). These include systematic classroom observations, analysis of classroom artifacts (e.g. student work, and ratings of teacher assignments), teaching portfolios, teacher self-reports of performance (e.g. surveys, EMA, teaching logs, and interview), and student ratings of teacher performance. This information is used to help support timely and efficient personnel decisions, including targeting teachers for improvement, termination, retention, or promotion. For the purposes of this review, which link to the data available for analysis in this study, the focus is on classroom observations, self-reports of practice, and EMA are the topics that will be covered.

**Classroom observations.** Classroom observations are the most widely used measure of teacher instructional quality. Procedures vary widely in what they evaluate and how they are conducted (Praetorius, Lenske, & Helmke, 2012). At most schools, systematic observations are

conducted by either a school administrator (e.g. principal) or an outside evaluator several times throughout the year. Often observation protocols looking at TE evaluate both general or subject specific teaching practices (Waxman et al., 2004). Observations typically address classroom processes including holistic aspects of instruction, interactions between teachers and students, as well as teacher practices (Pianta & Hamre, 2009). Additional observational foci may be measuring subject or context specific aspects of teaching. Observations are widely regarded as a useful source of feedback for professional development purposes (Pianta & Hamre, 2009). Although they might be good at providing feedback to an individual teacher, observational protocols used for teacher feedback and professional development vary widely. Standardized observational measures of teaching quality typically target aspects of TE such as classroom management, instructional strategies, teacher-student interactions, and presentation of subject matter knowledge (Goe et al., 2008). These domains are typically rated on a Likert-type scale. Overall, the research generally suggests that rigorously developed observational protocols are linked to student achievement, though associations are sometimes modest (Gallagher, 2004; Kimball et al., 2004; Milanowski et al., 2004).

Two observation protocols that are frequently used in research and by school districts to measure overall instructional quality are the University of Virginia's Classroom Assessment Scoring System (CLASS) and Charlotte Danielson's (1996) *Enhancing Professional Practice: A Framework for Teaching* (Pianta, La Paro, & Hamre, 2006). Both observational protocols have specific instructions and procedures for training raters and thus for establishing reliable scores. Among studies that have utilized different or modified versions of the aforementioned observational protocols, the research has been mixed. Research outcomes and validity assessments are very dependent on the sampling procedures, instruments used, and training of

raters (Praetorius et al., 2012). While observational protocols are widely used by schools as part of comprehensive evaluations, there is still a dearth of reliability and validity research on their use (Pianta & Hamre, 2009). The logistical barriers to using systematic observations systems are daunting. Major barriers are occur with the development and validation of these tools. Even when there is a valid and reliable protocol, it can be expensive to implement with high quality due to the cost of observers time, training and calibrating (Goe et al., 2008). Furthermore, purely observational methods may miss some important information, as they are not useful for assessing beliefs, feelings, intentions, or out of classroom activities (e.g. planning).

**Teacher surveys.** Teacher reports of what they are doing in classrooms is typically assessed through surveys, instructional logs, and interviews (Little et al., 2009). Survey measures may focus on overarching aspects of teaching important in all contexts or specific subject matter, grade levels, or content areas (Desimone, Smith, & Frisvold, 2010). They can consist of rating scales assessing the extent to which certain practices are used, reports of the frequency of certain practices or standards are implemented, or a checklist of easily observable behaviors (Goe et al., 2008). Large-scale surveys typically measure enactment of curriculum or reform-oriented instructional practices (Desimone et al., 2010).

Some examples of well-known surveys utilized by researchers and districts include California Learning Assessment System, Reform-Up-Close, Surveys of Enacted Curriculum, and Trends in International Mathematics and Science Study (TIMSS) (Little et al., 2009; Porter, 2002). These well-known teacher surveys address coverage of material, pedagogy, instructional materials and technology, and professional development (Little et al., 2009; Mullens, 1995). In general, there have been mixed results in regards to the reliability and validity of these instruments (Mayer, 1999; Porter, Kirst, Osthoff, Smithson, & Schneider, 1993).

Teacher self-report measures have certain advantages including the ability to tap into internal states (e.g. beliefs, intentions, knowledge, etc.), they are generally unobtrusive, cost-efficient, and ideal for gathering large amount of data at once (Desimone et al., 2010; Kihlstrom et al., 2000). However, as Goe et al. (2008) notes, “extreme caution should be taken not to base potentially consequential decisions on results of self-report measures” as research findings on many of these survey of instructional quality have yielded mixed results as to their reliability and validity (p. 30). In general, self-report responses are susceptible to social desirability, which may result in over or underreporting of certain practices (McIntyre et al., 2016).

**Ecological Momentary Assessment (EMA).** Much of the research on teachers and teaching has utilized retrospective and cross-sectional methodology (Klassen et al., 2011). This is a significant issue with the body of literature to-date in that the primary sources of data collection utilize observations, surveys, and interviews that are susceptible to retrieval distortion, psychological states, and the natural environment (McIntyre et al., 2016). One way to combat these issues is by using momentary procedures, or EMA, to obtain immediate reports and temporal patterns of activities in the classroom (Bolger, Davis, & Rafaeli, 2003; Holsti & Barr, 2006; Moskowitz & Young, 2006; Shiffman, 2000). EMA is an effective way to examine what teachers think, feel, and do over time and can be used as an alternative methodology to surveys (Carson, Weiss, & Templin, 2010). EMA is particularly suited to the study of teachers’ affective states and behaviors within their work environment which is highly changeable and complex (McIntyre et al., 2016). It is ecologically valid since it captures naturally occurring behavioral, physiological, social, and/or psychological processes in a real-time real-world setting (Stone, Shiffman, & DeVries, 1999).



Compared to surveys and observation, using EMA in studies offers some unique benefits. First, EMA is resistant to recall biases that are common in self-report measures (McIntyre et al., 2016; Moskowitz & Young, 2006). It has consistently been reported that individuals' responses in self-report measures can often be an over- or underestimation of their previous recollection of emotions, behaviors, and cognitions (Diener, Smith, & Fujita, 1995; Robinson & Clore, 2002; Shiffman, 2000; Stone et al., 1998; Thomas & Diener, 1990). Furthermore, when assessing these prior recollections they can often be contaminated by recency effects, mood state, and prevailing salient events (Brief, Butcher, & Roberson, 1995; Stone, Shiffman, & Devries, 1999).

A second benefit of EMA is that it is able to detect meaningful within-person variability that typically gets lost in aggregate-level data (Beil & Weiss, 2003). Thus, rather than conventional one-time measures, EMA allows researchers to address how phenomena progress and change over time (Carson, 2010). Finally, EMA also allows for the study of situational variables of psychological phenomena such as emotion. This allows researchers to better understand how incidences of positive and negative affective states result from various proximal stimuli, perceptions, and personality constructs (Alliger & Williams; Fisher, 2000, 2002; Fisher & Noble, 2004; Grandey, Tam, & Brauburger, 2002; Scollon, Kim-Prieto, & Diener, 2003; Zohar, Tzischinski, & Epstein, 2003).

EMA has been utilized to study daily experiences in a variety of disciplines including mental health, workplace behavior, healthy aging, alcoholism, smoking cessation, exercise habits, and interpersonal relationships (Fisher & Noble, 2004; Hausenblas et al., 2008; Havermans, Nicolson, & Devries, 2007; Graham, 2008; Husky, et al., 2007; Lee & Waite, 2005; Litt, Cooney, & Morse, 1998; Riediger & Freund, 2008; Scharf, Dunbar, & Shiffman, 2008;

Teuchmann, Totterdell, & Parker, 1999). However, there are very few studies on utilization of EMA for collecting data about teacher's daily experiences (McIntyre, 2016).

In one of the few studies of EMA with teachers, Aili and Brante (2007) employed EMA to determine the number of weekly working hours that teachers reported applying professional knowledge or performing work that qualifies as professional autonomy. Studying a small sample of teachers, Daniels, Hartley, and Travers (2006), found that teachers' pre-existing beliefs concerning the sources of stressful events on concurrent negative feelings impacted the actual relationship between on-the-job stressors and negative feelings. Although the findings from both of these studies were not focused on measuring teaching quality, these initial efforts document EMA's capability to assess detailed dynamic information about teachers within their professional environment.

While EMA is a promising means of measuring teacher emotions and behaviors in the classroom, there are several important strategic decisions when using this method such as careful consideration of design (e.g. multiple choice or slider response, software package, and framing of questions) and sampling issues (e.g. when, how often, and how long data should be collected) (Beal & Weiss, 2003; Reis & Gable, 2000; Scollon, Kim-Prieto, & Diener, 2003). Additionally, implementation of EMA data collection can be demanding on the respondents if the sampling is frequent, lengthy, inconvenient, or tedious. Participants may be interrupted multiple times during the day across several days and they may be expected to perform other tasks before, after, and throughout the period of data collection (Carson, 2010; McIntyre, 2016).

To summarize, EMA is an ecologically valid means of collecting data that allows teachers to report on their experiences repeatedly over time and across context within their natural environments. Since data are collected in real time in actual performance settings, they

can be generalized to real-life and real-world experiences more readily than other survey methods. When combined with data from observations, as well as surveys of constructs such as motivation, EMA can present a comprehensive picture of contextual variables, affective states, and behaviors that co-occur and affect teachers' general functioning in the classroom. Thus, using EMA procedures can articulate more clearly the various inputs and processes that lend themselves to high quality teaching and classroom practices.

### **Classroom Management**

To ensure student success, classroom teachers must deliver quality instruction, monitor student progress, and differentiate instruction (Stronge, 2007, 2011). However, a teacher's ability to effectively teach, monitor, and differentiate rests on their ability to manage student behavior in the classroom (Gest & Gest, 2005). Classroom management is broadly defined as the actions taken that are necessary to create and maintain an environment in which instruction and learning can occur, and supports instructional goals (Brophy, 1988; Emmer & Stough, 2001; LePage et al., 2005). Classroom management has historically been cited as the most challenging aspect of teaching for novice teachers (Carr, 2013; Hicks, 2012; Stronge et al., 2003). Poor classroom management has increasingly been seen as a concern given that it is a major cause of teacher attrition within the first five years (Hicks, 2012). Unfortunately, many teachers do not receive adequate training in classroom management techniques prior to beginning their teaching careers (Oliver & Reschly, 2007).

Effective classroom management includes implementing classroom rules and routines, structuring the classroom environment, actively monitoring students, using behavior reduction strategies, collecting and using data to monitor student behavior, and modifying classroom management procedures as needed (Oliver & Reschly, 2007). The use of rules and routines by

teachers serves as a preventative and powerful component of classroom management (Colvin, Kameenui, & Sugai, 1993; Kerr & Nelson, 2002; Oliver & Reschly, 2007). Additionally, effective teachers establish routines that allow for efficient use of time and fewer disruptions (Stage & Quiroz, 1997). After rules and routines have been established, effective teachers implement strategies to encourage students to follow these rules through such techniques as specific contingent praise, token economies, behavior contracts, etc. (Higgins, Williams, & McLaughlin, 2001; Marzano et al., 2003; Partin, Robertson, Maggin, Oliver, & Wehby, 2009; Stage & Quiroz, 1997).

Important aspects of effective structuring of the classroom environment include creating a physical arrangement that minimizes distractions, provides access to students, and allows for ease of traffic flow, plus allows for efficient use of class time, communicating appropriate behaviors for particular classroom activities, and ensuring positive student interactions (Guardino & Fullerton, 2010; Oliver & Reschly, 2007; Paine, 1983). Effective classroom managers also monitor the classroom regularly in order to be aware of what is happening in their classroom and to potentially detect inappropriate behavior early on (Good & Lavigne, 2017; Marzano et al., 2003). By doing so, teachers can potentially intervene to correct problem behavior before it escalates. These classroom management strategies are used by the teacher to increase students' cooperation and engagement and to decrease students' disruptive behaviors, thus facilitating a positive learning environment with more time devoted to instruction. Instructional time lost to disruptive behavior within the classroom has been shown to lead to lower academic engagement and achievement (DeJarnette & Sudeck, 2015). Thus, good classroom management can be viewed as an essential precursor for student learning (Emmer & Stough, 2001).

**Classroom management and teacher quality.** Effective classroom management has consistently been shown to improve student achievement by decreasing the incidence of disruptive behaviors, increasing student engagement, and enhancing the use of instructional time (Doyle, 1986; Wang et al., 1993). Research over the past 30 years strongly supports the notion that effective teachers practice effective classroom management and organizational behaviors (Becker, Madsen Jr, Arnold, & Thomas, 1967; Brophy, 1988; Emmer, Evertson, Clements, & Worsham, 1997; Emmer & Sabornie, 2013; Emmer & Hickman, 1991; Evertson & Neal, 2006; Evertson & Weinstein, 2013; Hattie, 2008; Porter & Brophy, 1988). In fact, having high quality classroom management skills is a robust predictor of TE and student learning gains (Baumert & Kunter, 2013; Stronge et al., 2011). For example, in a meta-analysis conducted by Marzano et al. (2003), the authors found that in classes with teachers who practiced effective classroom management techniques there was a 23 percentile point increase in engagement rates ( $SD = .617$ ) when compared to the engagement rates of poor classroom managers. Furthermore, among classes with effective classroom managers there was a 20 percentile point increase in student achievement ( $SD = .521$ ) when compared to teacher who were poor classroom managers.

In a study by Freiberg et al. (2009), the authors examined the effects of an instructional management and prosocial classroom intervention on reading and mathematics achievement data over two years. Using a stratified random sample of 350 elementary students, they found that students who received the intervention ranked at the 67<sup>th</sup> percentile whereas the control students ranked at the 50<sup>th</sup> percentile:  $F(1, 698) = 20.18, p < .001$ . Similarly in reading, the intervention group ranked at the 64<sup>th</sup> percentile while the control group ranked at the 50<sup>th</sup> percentile:  $F(1, 698) = 20.11, p < .0001$ . Additionally, the authors conducted analyses to determine the effect size between the intervention and control schools. They found when comparing the average

improvement there was a strong effect size for the intervention group in reading ( $E = .54$ ) and a medium effect size for the intervention group in math ( $E = .39$ ). The authors concluded that their study supported the notion that changes within the learning environment, particularly in improving classroom management had a positive effect on student achievement.

To summarize, classroom management is a precursor to student achievement and is considered one of the most important instructional skills for teachers to acquire (Stoughton, 2007; Stronge, 2007). Teachers who are effective classroom managers utilize strategies that increase student engagement and cooperation and decrease disruptive behaviors. As a result, they are able to provide an appropriate learning environment to foster student learning and engagement. Along with achievement and time-on task, classroom management skills have been associated with some internal motivational states, in particular self-efficacy (Conner et al., 2014; Hamre & Pianta, 2005; Nye et al., 2004).

### **Teacher Self-Efficacy**

Bandura's social cognitive theory provides a basis for understanding human behavior and motivation through the interactions of environmental, internal personal (e.g. cognitive, affective, and biological events), and behavioral factors (Bandura, 1977, 1986, 1997; Wood & Bandura, 1989). This theory assumes that humans are more than merely passive reactors to their environments; they are active shapers who are able to meet and affect the demands of their environment through adaptive actions. A central determinant of individuals' decisions to engage and persevere in certain behaviors is the individual's perceived self-efficacy.

Perceived self-efficacy describes a person's self-conception of their own performance capability in one or more distinct areas of functioning. One of the basic principles of self-efficacy theory is that people are more likely to engage in activities in which they have high self-

efficacy and less likely to engage in those they do not. Performance and motivation are in part determined by how effective people believe they can be. In fact, Bandura (1997) suggested that an over-inflated self-efficacy, or overestimation of one's abilities, is useful in that it can lead to greater effort and perseverance. Thus, highly efficacious people tend to be more effortful, invest more time, set higher goals, persist longer in the face of obstacles, and are more resilient when faced with setbacks.

Perceived self-efficacy is not an innate construct within an individual but rather one that builds, adapts and changes based on a variety of environmental sources. Bandura (1977, 1997) identified four of these sources which include the following: 1) mastery experiences, 2) vicarious experiences, 3) verbal persuasion, and 4) physiological feedback. Mastery experiences refer to self-efficacy information that is gained via an individual's performance on a particular task. When an individual perceives that their performance has been successful this experience can raise efficacy beliefs, as well as provide a foundation for their beliefs regarding future performance in similar contexts (Cantrell, Young, & Moore, 2003). Vicarious experiences are another source of self-efficacy information that is gained via observation of models. Whether or not one gains self-efficacy information from this experience depends largely on degree to which the individual identifies with the model (Bandura, 1977). Verbal persuasion refers to efficacy information gained from both positive and/or negative feedback about an individual's capability to perform a particular task from a credible source (Bandura, 1997). Finally, physiological and affective states refers to efficacy information gained through an individual's physiological and emotional reactions to a particular task. Of the four sources, mastery experiences are considered to be the most powerful source of efficacy information as it provides immediate evidence of an individual's performance in a given situation. Success or failure may lead to increased or

decreased self-efficacy (Tschannen-Moran et al., 1998). These performance-based accomplishments or failures are considered the most powerful agent of change to one's perceived self-efficacy.

Self-efficacy research has extended to a variety of disciplines including health behaviors, job performance, resilience, and education (Benight & Bandura, 2004; Holden, Moncher, Schinke, & Barker, 1990; Moritz, Feltz, Fahrback, & Mack, 2000; Multon, Brown, & Lent, 1991; Stajkovic & Luthans, 1998). Although not as well developed as some other areas, there is a growing body of research on TSE and how it relates to teacher quality, student achievement, and teacher attrition and burnout. TSE represents teachers' belief that they are able to perform those teaching behaviors that bring about student learning even when facing difficult or unmotivated students (Klassen et al., 2011; Ross, 1998; Wolters & Daugherty, 2007). Teachers with high self-efficacy tend to utilize high levels of instructional strategies, student engagement, and classroom management strategies, all of which constitute quality teaching practices (Allinder, 1994; Tschannen-Moran & Hoy, 2007; Zee & Koomen, 2016).

TSE was first introduced by the RAND Corporation in the late 1970s. RAND investigated innovative educational programs funded by the federal government (Ashton & Webb, 1986). RAND proposed that TSE beliefs, or teachers' self-referent judgments of capability, are relevant for a range of outcomes at different levels of classroom ecology. TSE is thought to impact teacher quality such that teachers with an assured sense of self-efficacy set the tone for a high-quality classroom environment by planning lessons that advance students' abilities, making efforts to involve them in meaningful ways, and effectively managing student misbehavior (Allinder, 1994; Chacon, 2005; Woolfolk et al., 1990).



**Theory and measurement of teacher self-efficacy.** The origins of TSE have been historically derived from two theories; Rotter (1966) locus of control and Bandura (1977) social cognitive theory. Both theories emphasize human agency or the idea that individuals are able to exercise control over their actions that affect their lives (Bandura, 1997). Rotter (1966) conceptualized locus of control as a generalized expectancy for control of reinforcement that individuals develop in relation to their environment. Internal control means that an individual's perception of control derives from within the person, external control means that the individual perceives control to be outside the person, possibly in powerful others, in objective external conditions, or in chance or luck. According to Rotter (1966), people can differ in their perceptions of whether outcomes are contingent on luck, fate, or others (external control), or a result of their own actions (internal locus). These different perspectives are considered to be largely determined by an individual's interaction with their environment that reinforce an individual's actions.

Those who believe or expect that they can control their own destinies behave differently than those who expect that their outcomes are outside of their control. In general, those who believe their environment is responsive to their actions (internal locus of control) tend to have higher well-being and success (Lachman, 2006). Rotter's (1966) theory laid the groundwork for many studies and scales, including the first measure of TSE in the 1970's. The first TSE measure was utilized by RAND researchers in two studies, one on teacher success in teaching reading to minority students (Armor et al.) and an evaluation of teachers' use of federally funded innovations (Berman, MacLaughlin, & Bass, 1977). Both studies utilized a two item scale tapping into teachers' perceptions of their own capabilities. Using Rotter's locus of control they defined TSE as the extent to which teachers believed that they could control the reinforcement of

their actions, that is, whether control of reinforcement was internal or external. The authors found that the extent to which a teacher believed that the consequences of teaching (i.e. student motivation and learning) were in the hands of the teacher rather than externally controlled predicted student success on standardized reading assessments.

In the second study, RAND researchers found TSE to be a strong predictor of continuation of innovative instructional practices even after federal funding was withdrawn (Berman et al., 1977). The success of the RAND studies propelled many other researchers to investigate and expand upon TSE. However, many were skeptical about the reliability of the two-item scale and attempted to develop longer and more reliable measures. Among these were Guskey (1982), responsibility for student achievement (RSE), Rose and Medway (1981) teacher locus of control (TLC), and Ashton and Webb (1982) Webb efficacy scale.

The Rand Corporation's investigation of TSE also sparked a second conceptual strand around the same time. This work on TSE grew out of Bandura's (1977) social cognitive theory. Bandura's (1997) concept of self-efficacy was defined as "the beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (p. 3). Bandura further added to this theory by making a distinction between outcome expectancies and self-efficacy expectations. Outcome expectancies were conceptualized as an individual's estimate of the likely consequences of performing a task at their expected level of competence and were generally considered to be operationally equivalent to Rotter's (1966) construct. In general this construct determined whether the social environment is perceived to be reactive to personal actions or not.

Self-efficacy expectations refer to an individual's conviction that he or she can orchestrate the necessary actions to perform a given task, which goes beyond environmental

contingencies. Bandura (1997) argued that whether or not a person knows that an action will result in desired outcome is inconsequential when they lack the beliefs in their ability to produce such actions. For example, if a teacher does not believe they have the ability to perform a teaching task they are unlikely to even initiate a teaching task. Thus, he emphasized that self-efficacy expectations, not only predict outcome expectancies, but were far more useful in predicting human behavior. Bandura (1997) described self-efficacy beliefs as influencing thought patterns and emotions. These thought patterns, in turn, enable actions in which people expend substantial effort and persistence in pursuit of goals, rebound from temporary setbacks, and exercise some control over events that affect their lives (Bandura, 1986, 1993, 1997).

Bandura's (1997) conceptualization of self-efficacy greatly impacted TSE research. One of the major contributions was that Bandura (1977, 1997) was the first to differentiate between self-efficacy and outcome expectancies. In work by Gibson and Dembo (1984) in their development of a TSE measure, the authors found modest evidence for two independent factors that greatly resembled self-efficacy and response-outcome expectancies. These two factors supported the two expectancies of Bandura's social cognitive theory. They labeled these factors as personal teaching efficacy (PTE,  $\alpha = 0.75$ ) which reflected self-efficacy, and general teaching efficacy (GTE,  $\alpha = 0.79$ ) which they assumed captured outcome expectancy. Many researchers, using Gibson and Dembo's items confirmed the existence of these two factors (Anderson et al., 1988; Burley et al., 1991, Hoy & Woolfolk, 1993; Moore & Esselman, 1992; Soodak & Podell, 1993). Gibson and Dembo's measure was the most popular and is prolific in studies of the TSE from the late 1980s to the late 1990s. However, its popularity has since declined due to issues with content and construct validity of the factors (Pajares, 1997; Tschannen-Moran & Woolfolk Hoy, 2001; Woolfolk & Hoy, 1990).

An additional consequence of Bandura's social cognitive framing of self-efficacy is that researchers began to conceptualize TSE as one that is task- rather than situation-specific (Tschannen-Moran & Woolfolk Hoy, 2001; Woolfolk & Hoy, 1990). Bandura (1997) stated that TSE should not be measured as an omnibus trait given the many different types of tasks teachers must perform, different subjects, and different grades. Further, he argued that the benefit of this conceptualization of TSE was that it would have higher predictive validity, due to variations in TSE across different tasks (1997). Thus, Bandura created a 30-item measure that included seven subscales: (1) efficacy to influence decision making, (2) instructional efficacy, (3) efficacy to influence school resources, (4) efficacy to create a positive school climate, (5) efficacy to enlist community involvement, (6) efficacy to enlist parental involvement, and (7) disciplinary efficacy. Bandura's measure was designed to provide a multifaceted picture of teachers' efficacy beliefs without becoming too narrow or specific, however, no reliability or validity information has been available. Bandura's measure of TSE was not used outside his own personal research; however, since his new conceptualization of TSE, multiple measures have since been developed ranging from very general to specific (Tschannen-Moran & Woolfolk Hoy, 2001).

Conceptualizing TSE as task-specific presents researchers with an important issue: finding the optimal level of specificity of the measure. The most popular of these measures is Tschannen-Moran and Woolfolk Hoy Teachers' Sense of Efficacy Scale (TSES) (Tschannen-Moran & Woolfolk Hoy, 2001). This measure is a task specific instrument comprised of three dimensions: (1) efficacy in instructional practices, (2) efficacy in classroom management, and (3) student engagement. This measure has been described as "superior to previous measures of teacher efficacy in that it has a unified and stable factor structure" (Woolfolk Hoy & Burke Spero, 2005, p. 354). Other investigators have attested to its satisfactory reliability and construct

validity across grades and subject areas (Klassen et al., 2009; Tschannen-Moran & Woolfolk Hoy, 2001). However, some have noted that having only three task-specific domains is inadequate and that most items lack clear obstacles, which is strongly recommended by Bandura (1997) (Skaalvik & Skaalvik, 2007).

In addition to the dimensional approach of measuring TSE, other researchers have developed measures that are subject-specific. This includes scales for literacy, inclusive practices, science, discipline, and use of technology (Brouwers & Tomic, 2000; Malinen et al., 2013; Riggs & Enoch, 1990; Sang, Valcke van Braak, & Tondeur, 2010; Tschannen-Moran & Johnson, 2011). Taken together these studies with varying types and levels of specificity show that TSE is reflected in multiple specific components of the teachers' profession and that the strength of TSE can fluctuate between teaching tasks, roles, students, and over time.

**Correlates, antecedents and consequences of TSE.** TSE is an outcome of interactions of environmental, internal personal (e.g. cognitive, affective, and biological events), and behavioral factors. While TSE, once developed, is considered to be relatively stable over time, it can be subject to change. As TSE is generative it influences goal setting and persistence, and in doing so contributes to mastery experiences that can heighten or depress efficacy further (Bandura, 1997). Efficacy beliefs, while mostly stable over time are most pliable early on in learning. As previously stated, mastery experiences are one of the strongest influences on self-efficacy and these experiences can heighten or depress efficacy further (Tschannen-Moran et al., 1998). Because new teachers may lack the knowledge or skills to attain mastery experiences, early on in a teachers' career the other sources of influence on TSE may take a more meaningful role in shaping TSE, nonetheless, mastery experiences remain the most potent (Burley, Wade, Villeme, & Brockmeier, 1991; Cantrell, Young, & Moore, 2003; Caprara, Vittorio, 2006). There

is a general consensus in the literature of significant relations between years of experience in teaching and TSE (Caprara, 2006; Chong, 2010; Dembo, 1985; Dicke, 2014; Kerns, 2015). Novice teachers have been reported to have a decline in TSE during the first two years of their career (Ross, 1996). In the United States, up to 25% of beginning teachers do not return for their third year and almost 40% leave before five years of teaching (Goldring et al., 2014; Tschannen-Moran & Hoy, 2007). In general, those who tend to leave their teaching careers have significantly lower TSE than teachers who remain in teaching (Aloe, 2014; Dicke, 2014; Wang, 2015). TSE has also been linked to the amount of stress experienced by teachers, and level of professional commitment (Coladarci, 1992; Evans & Tribble, 1986; Tschannen-Moran & Hoy, 2007). In a particular study by Tschannen-Moran and Hoy (2007), they found that more experienced teachers had higher TSE overall, as well as higher efficacy in two of the three subscales of TSES; efficacy for instructional strategies and efficacy for classroom management. In general, demographic variables such as race, gender have not been significant predictors of TSE beliefs for preservice or experienced teachers (Tschannen-Moran & Hoy, 2007).

**TSE and instructional quality.** Studies on TSE suggest that instructional behaviors, practices, and strategies that teachers employ in the classroom to encourage students cognitive development may, in part, be determined by their self-efficacy (Chacon, 2005; Tschannen-Moran & Hoy, 2007; Tschannen-Moran et al., 1998). In fact, many cross-sectional studies have shown positive relations between TSE and student motivation and academic achievement (Caprara et al., 2006; Midgley, Feldlaufer, & Eccles, 1989; Ross, 1992; Thoonen, Slegers, Oort, Peetsma, & Geijsel, 2011). TSE has also been investigated in its relations to instructional quality overall and specific instructional behaviors (Tschannen-Moran et al., 1998; Zee & Koomen, 2016). Some of the most notable relations between levels of TSE and various instructional practices;

teachers who have high TSE tend to include more use of innovative teaching methods, practice more effective classroom management, utilize more diverse instructional strategies, establish positive classroom climates, and are successful in teaching students with diverse learning needs (Allinder, 1994; Caprara et al., 2006; Coladarci & Breton, 1997; Cornelius-White, 2007; Dembo & Gibson, 1985; Dicke et al., 2014; Guskey, 1988; Soodak & Podell, 1993). However, the assumption of a positive causal relation between TSE and instructional behaviors and student achievement is based almost exclusively on cross-sectional data.

In a comprehensive literature review of TSE and classroom level outcomes, Zee and Koomen (2016), included 25 studies in their meta-analysis examining the effects of TSE on classroom processes, achievement, and teacher well-being. Unfortunately, of these studies, less than half utilized samples above 50 teachers and approximately 70% of the articles they reviewed used cross sectional designs, simple correlations and global measures of TSE. Unsurprisingly, it was found that teachers' overall levels of instructional support was not found to be affected by TSE, especially for preservice teachers (Guo, Piasta, Justice, Kadaverek, 2010; Pakarinen et al., 2010). Among inservice teachers it was found that TSE contributed to a variety of instructional practices including process-oriented instruction and differentiation, ability to connect to students' lives, the number of goal changes made, the ability to employ effective teaching strategies, and utilizing instructional practices supporting inclusive education (Allinder, 1995; Martin, Sass, & Schmitt, 2012; Thoonen, Slegers, Oort, Peetsma, & Geijsel, 2011; Weshah, 2012). In a study employing structural equation modeling by Geijsel, Slegers, Stoel, and Kruger (2009), the authors utilized a sizable sample of teachers ( $N = 328$ ) to investigate the importance of teachers' school organizational conditions, leadership practices of principals, and psychological states on participation in professional learning activities. The authors found strong

relations between a teachers' sense of self-efficacy and their participation in professional learning activities. In particular, they found that teachers with high TSE were more likely to try out new instructional approaches to improve their teaching ( $r = .30$ ), keep up to date on professional practices ( $r = .31$ ), and change their practice to promote process-oriented student learning ( $r = .25$ ).

There have also been studies relating TSE to instructional support in specific subjects, most notably math and literacy (Guo et al., 2010; Holzberger et al., 2013; Midgley et al., 1989; Pajares & Miller, 1994). Several small-scale correlational studies have been conducted on relations between TSE and literacy. These studies generally found that while teaching students to read, TSE had no relation to the number of reading strategies utilized although it did relate to preservice teachers use of expository text as a reading tool (Haverback, 2009; Yildrin & Ates, 2012). Among inservice teachers, however, TSE was found to contribute to both the quality of the literacy instructional support and the instructional literacy environment in preschool (Guo, Sawyer, Justice, & Kaderavek, 2013; Justice, Mashburn, Hamre, & Pianta, 2008). Regarding mathematics development, the small amount of literature suggests little to no relations between TSE and observed mathematics instructional practices, individual learning support in math, or level of student cognitive activation. Surprisingly, a study by Holzberger, Philipp, and Kunter (2013), showed a reverse effect of instructional quality on TSE. Using a sample of 155 German secondary mathematics teachers and 3,483 students they collected data at two points from self-report measures of TSE, and teacher and student ratings of instructional quality (i.e. classroom management, cognitive activation, and individual learning supports). Using structural equation modeling they found that teachers' ratings of classroom management and students' experience of cognitive activation predicted teachers' subsequent self-efficacy. Cross sectional analyses



revealed that teachers with higher self-efficacy beliefs showed higher instructional quality by both teacher and student raters. Larger effects sizes were found for teacher ratings suggesting a common method bias. Further analyses were conducted using a cross-lagged panel design over two points of time one year apart. Surprisingly they found that when teachers' baseline instructional quality was controlled for there were small, nonsignificant effects of TSE on later instructional quality.

**TSE and classroom management.** There is an emerging body of literature on classroom management and its relation to self-efficacy. Teachers' effective in classroom management must possess knowledge, skills, and a belief in their capability to be proactive and reactive in order to maintain an environment conducive to educational goals. According to Bandura (1986), self-belief is a requirement for personal accomplishment and as individuals successfully accomplish goals or tasks, they build a belief in their own personal efficacy. Thus, if a teacher does not believe in their capability to managing the classroom effectively they will be less likely to act as self-doubts often trump both knowledge and skill (Bandura, 1997). Studies on preservice teachers in particular, demonstrate that TSE has a beneficial impact on lesson presenting and classroom management behaviors (Saklofske, Michayluk, & Randhawa, 1988). Among inservice teachers TSE has also been shown to contribute to their use of preventative approaches of managing student discipline concerns and overall classroom management strategies (Abu-Tineh et al., 2011; Morris-Rothschild & Brassard, 2006). Among preservice and inservice teachers alike, those who report lower levels of TSE tend to report higher stress as a result of discipline issues (Klassen & Chiu, 2010, 2011).

Classroom self-efficacy (CMSE), or the extent to which a teacher feels that he or she is competent in keeping students engaged and on-task, organizing the classroom, and maintaining

order, has received increasing attention within TSE literature (Emmer, 1990; Emmer & Hickman, 1991). Measures of TSE with more than one domain typically have detected classroom management as a separate and valid factor (O'Neill & Stephenson, 2011; Skaalvik & Skaalvik, 2007; Tschannen-Moran & Woolfolk Hoy, 1998). For example, a number of studies have investigated the subscale of CMSE from common TSE instruments (Aloe, Amo, & Shanahan, 2013; O'Neill & Stephenson) or constructed their own CMSE instruments by combining items from multiple TSE scales (Change, 2009; Ho & Hau, 2004; Ozdemir, 2007).

Classroom management and TSE have primarily been studied within the context of teacher burnout, stress, and attrition (Brouwers & Tomic, 1998, 2000; Leiter, 1992; Cherniss, 1993; Chwalsz, Altmaier, & Russell, 1992; Friedman & Farber, 1992; Greenglass & Burke, 1988). Many studies have concluded that classroom management is positively associated with a sense of accomplishment in teachers and negatively associated with depersonalization and emotional exhaustion (Brouwers & Tomic, 2000; Chang, 2009; Evers, Tomic, & Brouwers, Friedman & Kass, 2003; Ransford, 2007; Skaalvik & Skaalvik, 2007). In a study by Fives, Hamman, and Olivarez (2007), the authors found that teachers who had high levels of self-efficacy for classroom management and instructional strategies were less likely to depersonalize their students and experience emotional exhaustion. Additionally, in a meta-analysis by Aloe et al. (2013), they found a moderate negative relationship between burnout and CMSE. However, others have argued that students discipline concerns are the most significant factor contributing to teacher burnout (Change, 2013; Friedman, 1995).

### **Gaps in the Literature & Aims of the Proposed Study**

There have been numerous studies on TSE and its relationship to a multitude of variables related to teacher quality. However, there are serious limitations to the literature, especially in

relating TSE to overall instructional quality and the potential mediating variables impacting this relationship (Aloe et al., 2013; Darling-Hammond, 2003; Klassen et al., 2010; Ross, 1998; Tschannen-Moran et al., 1998; Zee & Koomen, 2016). Furthermore, there are serious methodological issues in the body of research on TSE including, overabundance of cross-sectional studies, almost exclusive reliance on retrospective self-report measures of instructional quality, lack of control over teacher variables, and small sample sizes. Additionally, the existing body of literature has utilized a variety of different measures of TSE. This is problematic because these measures each conceptualize TSE differently making generalizations of relations between TSE and variables related to teacher quality impossible. Similarly, the conceptualization and measurement of teacher quality has varied widely in the literature (Corbett & Wilson, 2002; Goe, 2007; Goe et al., 2008; Hattie, 2008; Little et al., 2009). More recently, teacher quality has been limited to mean achievement gains derived from value-added measures of standardized test scores. While this is a proven measure that has utility in differentiating between high and low quality teachers, it cannot elucidate why some teachers produce greater learning gains in their students than others. A teachers' TE, or teaching practices both in and outside the classroom (i.e. lesson planning), is a measurable aspect of teacher quality that can reveal effective practices of quality teachers and further clarify why some teachers outperform others. Thus, the purpose of this study is to investigate the nature of the relationship between TSE and TE. More specifically, this study will investigate how teachers' classroom management mediates the relation between TSE and teachers' overall TE.

**Research questions and hypotheses.** The following research questions will be investigated through the use of teacher surveys, EMA data, and observational protocol ratings.

1. Does TSE directly predict teaching effectiveness?
  - a. It is hypothesized that TSE will directly predict both self-reported and observer-reported teaching effectiveness.
2. Does TSE directly predict classroom management?
  - a. It is hypothesized that TSE will directly predict both self-reported and observer-reported classroom management.
3. Does TSE have an indirect impact on teaching effectiveness via classroom management?
  - a. It is hypothesized that TSE will have an indirect impact on self-reported teaching effectiveness via self-reported classroom management.
  - b. It is hypothesized that TSE will have an indirect impact on self-reported teaching effectiveness via observer-reported classroom management.
  - c. It is hypothesized that TSE will have an indirect impact on observer-reported teaching effectiveness via self-reported classroom management.
  - d. It is hypothesized that TSE will have an indirect impact on observer-reported teaching effectiveness via observer-reported classroom management.

## **Chapter III:**

### **Methods**

This study was conducted using existing multi-method data collected for a large, multisite, longitudinal project designed to identify predictors and outcomes of job stress in urban middle school teachers. This is part of a larger study by McIntyre et al. (2016) and relevant procedures are described in detail in the methods section of their paper. The variables to be examined in this study include teaching effectiveness, TSE, and classroom management.

#### **Participants**

Participants in the proposed study included 202, 6<sup>th</sup>-8<sup>th</sup> grade teachers from 22 middle schools in two urban school districts in southern United States. Teachers were predominantly female (77%) with a mean age of 40.8 ( $SD = 11.2$ ), and mean number of years of experience was 12.5 years ( $SD = 9.1$ ). Teachers were 49.5% Caucasian, 33.2% African American, 10.4 % Hispanic, and 7% of other ethnicities. Appendix A provides further demographic information.

In order to determine the number of participants necessary to find significant effects, a power analysis was conducted using Proc Glimpower SAS 9.4 random Multireg option, with 202 teachers at  $p = .05$ , the study would be able to detect a partial regression coefficient of .2 with a power of  $\beta = .812$ . A regression weight of .2 falls between a small and medium effect size (Cohen, 1988).

#### **Study Design**

The teacher stress study was prospective and longitudinal. Data were collected in three waves over three school years from 2011-2013. The first wave took place over three consecutive school days between September and December. The second wave took place in the winter between January and March on one school day. Finally, the third wave took place in the Spring

at the end of the school year (April-May) over two consecutive days. Attrition over three years was low at 6.9% ( $n = 14$ ). Additions were made to the sample; in year one, three new teachers entered the study in winter, and during year two, 19 new teachers entered into the study to offset district turnover (turnover between year 1 and 2 was 15.1%).

## **Measures**

The teacher stress study used multi-method measures of malleable teacher factors such as stressors, affect, work attitudes, health, and classroom behavior. Data were collected on teachers through standardized self-report surveys, ecological momentary assessment (EMA), concurrent physiological measurements, and observational ratings. For the present study several measures were utilized from the existing data set including the socio-demographic survey, the Teacher Stress Diary (TSD), and the Texas Teacher Appraisal System (TTAS) and ratings of student behavior.

**Socio-demographic survey.** Teachers' demographic information came from a 31-item survey developed by McIntyre, McIntyre, and Durand (2010). This survey's items consisted of teacher demographics (e.g. ethnicity, age, subject(s) taught), professional characteristics (e.g. daily workload, tenure), and school characteristics (e.g. school enrollment, district).

**The Teacher Stress Diary (TSD).** McIntyre and McIntyre (2011) adapted the TSD, which is a validated iPod-based diary consisting of 61 short items that assessed momentary job conditions (e.g. control, demand), stress responses (e.g. cognitive responses, positive and negative emotions) resources (e.g. social support), student discipline issues, and overall perceived teaching effectiveness. Teachers entered responses that were either in a dichotomous or a visual analogue scale (0-100, "Not at all" to "Very") format. Scale scores were calculated by averaging scale items at each time point. Before entering their responses, teachers were asked to

indicate the activity they were currently engaged in using multiple choice items (e.g. conference, teaching, planning). The 61 TSD were only administered in full if the teacher had previously been engaged in teaching activities. If teachers had been previously engaged in planning or conference activities the 17 items on teaching effectiveness and student behavior were skipped. Entries were prompted by an alarm at the end of the class period to avoid interference with teaching activities. The time it took for teachers to complete these 61 items was between 2-10 minutes. The average scale reliability coefficients over six waves (two years) of data ranged from .82 to .98 (McIntyre et al., 2016).

Teachers' completed the TSD between six and eight times during the duration of the original study. The overall average of each of these points of data both at the between- (i.e. variability of individual teachers' average from the total sample of teachers) and within-teacher (i.e. variability of the average for each teacher amongst their different self-ratings) level was calculated and used in the present study as an ecologically valid measure of self-reported teaching effectiveness. Furthermore, questions related to classroom management were also averaged over time and used as an ecologically valid measure of self-reported classroom management effectiveness.

**Texas Teacher Appraisal System (TTAS) and ratings of student behavior.** The current study used a modified version of the TTAS to conduct classroom observations. In the current study, trained observers rated teachers on 66 items in four domains: Instruction, presentation, classroom management, and classroom environment. There were also seven items rating student problem behaviors and an overall rating of teaching effectiveness. All observers were trained prior to and following the first observation by the study's PI. Reliability

observations were conducted on 10% of the classroom observations by study co-investigators. Reliability ratings were in excess of .80.

Observers in the original study completed between six to eight classroom observations of each teacher. For the present study the average of each of these points of data both at the between and within-teacher level was calculated as a measure of both observer-reported teaching effectiveness and observer-reported classroom management effectiveness.

**Norwegian Teacher Self-Efficacy Scale (NTSES).** Skaalvik and Skaalvik (2007), NTSES is a 24-item measure of TSE that yields a standard score for overall TSE and its six subscales. These subscales include (1) efficacy for instruction, (2) efficacy for adapting education to individual students' needs, (3) efficacy for motivating students, (4) efficacy for keeping discipline, (5) efficacy for cooperating with colleagues and parents, and (6) efficacy for coping with changes and challenges. The authors originally developed their scale through an analysis of role expectations of teachers in Norwegian schools. Utilizing a sample of 2,249 Norwegian elementary school teachers, the authors found that the six factors explained 61% of the variance. They conducted a confirmatory factor analysis testing three models on the 24 items in the NTSES which confirmed the multidimensionality of the TSE construct while also showing that TSE can be analyzed as a latent trait based on the six subscales. The NTSES has been utilized internationally in studies to assess TSE including Norway, Canada, Italy and the United States (Avanzi et al., 2013; Skaalvik & Skaalvik, 2007; Wang, Hall, & Rahimi, 2015).

In the original study the NTSES was administered in the fall and winter of the first and second wave of data collection. Thus, teachers' NTSES scores could be collected up to four times. In the present study, the average of the NTSES scores both at the between and within-teacher level was calculated as a measure of their overall TSE.



## **Procedures**

In the original study, teachers were recruited through an informational session at their respective schools. Participation was voluntary. Teachers completed informed consent forms and baseline questionnaires assessing study variables. They also provided their class schedule. Each teacher received a one-time \$75 gift card for their participation in the study. Teachers also received a training session (10-15 minutes) on the TSD and data collection procedures. iPods were picked up by teachers prior to the first class period and subsequently returned at the end of the school day. For subsequent waves, a 2-3 minutes refresher training course on the TSD was conducted. The application was specifically programmed to the bell schedule of the schools and alerted the teachers five minutes prior to the bell as a reminder to complete the TSD questions.

## **Analysis Plan**

Hypotheses were evaluated using multilevel mediation models in Mplus 8.1. Data are repeated within teachers across time. Because data are nested within teacher, the study fit a random intercept for teachers. Additionally, because data are longitudinal, with some teachers being observed eight times, the study specified an autoregressive covariance structure. Autoregressive covariance structures assume that proximal time points are more highly correlated than distal time points. Specifically, the autoregressive covariance structure utilized included a time variable to indicate the duration between time points (two observations were in the fall and two in the spring and as such, the two observations in each season will be more highly correlated than the fall observations and the spring observations, per the autoregressive covariance structure). Study hypotheses were evaluated using the teacher level (level 2) regression parameters.

The following pathways were examined: (1) the direct relationship of TSE to observer and teacher ratings of classroom management, (2) the direct relationship of TSE to observer and teacher ratings of teacher quality, and (3) the direct relationship of observer and teacher ratings of classroom management and observer and teacher ratings of teacher quality. Descriptive statistics were analyzed to determine within- and between-teacher correlations and direct and indirect relations between the variables. The Sobel test was used to determine if mediators are significant. The Sobel test has been criticized for being overly conservative, and requiring very large sample sizes to achieve sufficient statistical power (Preacher & Leonardelli, 2001; Zhao, Lynch Jr, & Chen, 2010). Some have estimated that sample sizes greater than 500 are needed for mediator analysis using the Sobel test, and the recommended alternative analysis involves bootstrapping (Hayes, 2009). However, this recommendation is much simpler data structures, and is not feasible for this study. Furthermore, bootstrapping with this type of multi-level, longitudinal data is experimental, and programs for conducting these analysis are not available. Therefore, it was beyond the scope of this study to engage in bootstrapping to conduct the mediator analysis. As an alternative to the presumably conservative Sobel test, in this study non-significant Sobel tests for mediation would be followed-up with the Baron and Kenny method for inferring mediation (Hayes, 2009; Zhao et al., 2010). Although the Barron and Kenny method has some limitations, it is not nearly as power intensive as the Sobel test and provides some protection against Type II errors (Hayes, 2009; Preacher & Leonardelli, 2001; Zhao et al., 2010).

## Results

The following chapter is divided into two sections. The first section presents an overview of the descriptive statistics for each mediation model, as well as direct and indirect effects. The second section focuses on research hypotheses. This organizational structure is utilized because hypotheses focus on specific model paths in each of the four models.

### Descriptive Statistics and Direct and Indirect Relations

In each model figure the blue line in the center separates within- and between-teacher correlations or direct and indirect relations. Within-teacher or person-level deviations are denoted by  $e_N$  and fall below the dotted line. Between-teacher correlations are presented above the dotted line.

#### **Model #1: Observer-rated TE and observer-rated classroom management model.**

Table 1 presents the between (upper) and within (lower) teacher correlations, means, between and within standard deviations and interclass correlation coefficients (ICC) for the first model.

Table 1  
*Model #1 Correlations, Means, Standard Deviations, and ICC*

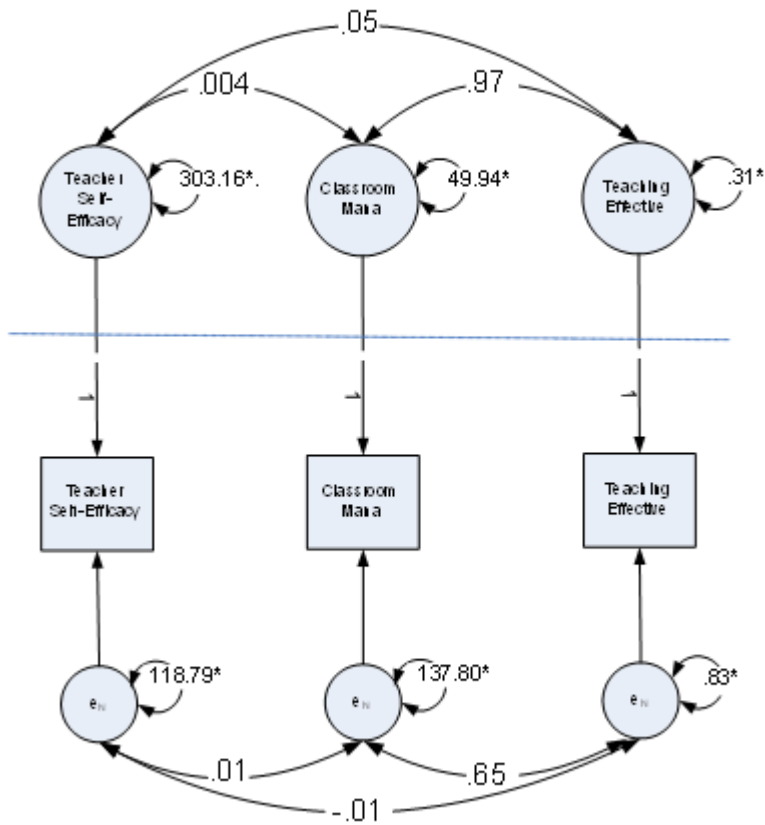
	TSE	CM	TEFF
TSE	1	0.004	0.05
CM	0.01	1	0.97*
TEFF	-0.01	0.65*	1
Mean	127.82	70.55	3.59
SD between	17.41	7.07	0.56
SD within	10.90	11.74	0.91
ICC	.72	.27	.27

The mean score among teachers on the NTSES, a measure of TSE, was 127.82, which is out of 168 total score, indicates that this sample of teachers tended to rate themselves as having above average TSE. In regards to observer-ratings of classroom management, there was a mean rating of 70.55 out of a total of 90 possible points. On average, observers rated most teachers as

being good classroom managers. The mean score for observer-rated teaching effectiveness (TE) was 3.59 out of a total of five possible points. In general, observers tended to rate teachers as slightly above average in their TE.

The interclass correlation coefficients (ICC) represents the proportion of variability at the teacher level and is an index of consistency of ratings within teachers. The ICC for TSE was .72 indicating that there is a great deal of consistency and stability between different teachers regarding efficacy suggesting that TSE is trait-like. However, compared to a maximum of eight measurement points for observer ratings of classroom management and TE, TSE was only measured four times during data collection and by the same rater. Thus, a portion of this within-teacher consistency on TSE could be an artifact of there being half as many ratings of TSE relative to observer-reported classroom management or observer-reported teaching efficacy. Furthermore, different observers rated different teachers at different points in data collection. Both classroom management and TE had ICCs of .27 suggesting that ratings among the observers of the same teacher varied considerably. Furthermore, it indicates that most of the variability in the data is occurring among observations of the same teacher rather than between teachers.

Figure 1. Model 1 Variable Correlations



Pearson correlations coefficients were computed to assess the relations between the variables in Model 1 Figure 1 displays these correlations, as well as the variance among the different variables. There was a very strong positive correlation between observer-reported classroom management and observer-reported TE at the between- and within teacher level. This indicates that, at the between-level, teachers who are rated highly on classroom management by observers tend to be the same teachers rated highly on teaching efficacy by observers,  $r = .97, p < .001$ . Also, at the within level, the class periods for a given teacher that were rated high on classroom management by observers also tended to be the class periods that were rated as having effective teaching by observers,  $r = .65, p < .001$ . The correlations between TSE and observer-

reported classroom management at the between- and within teacher level were nonsignificant ( $r_{\text{between}} = .004$ ;  $r_{\text{within}} = .01$ ). Furthermore, the correlations between TSE and observer-reported TE at the between and within levels were also nonsignificant.

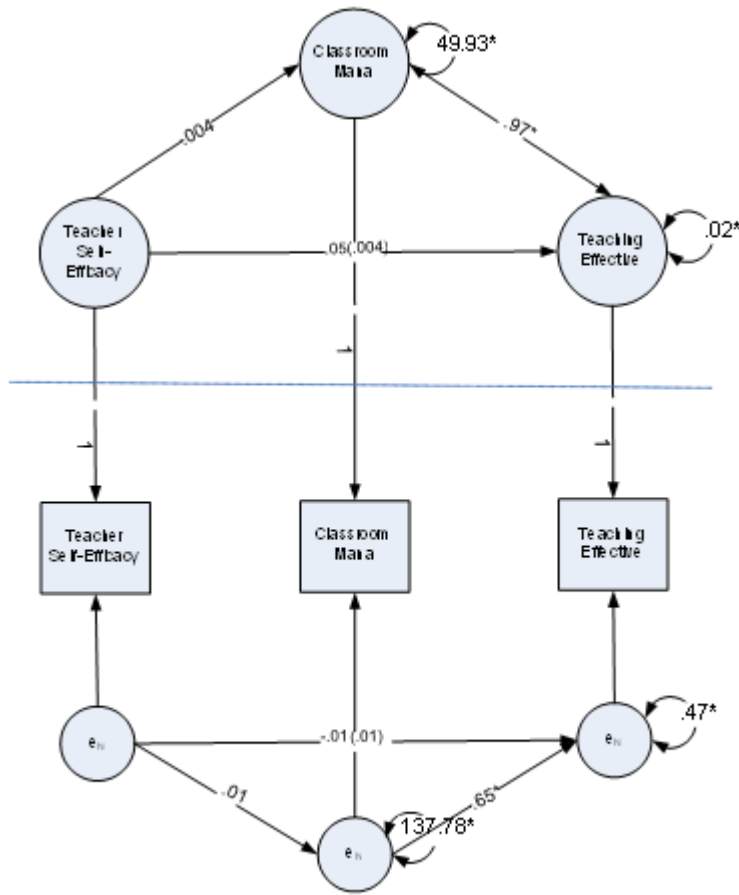
Table 2

*Model #1: Direct and Indirect Effects*

	<b>Outcome</b>	<b>Predictor</b>	<b>Relationship</b>	<b>Est.</b>	<b>P</b>
Between	TEFF	CM	Dir	0.97	<0.001***
		TSE	Dir	0.05	0.36
	CM	TSE	Dir	0.004	0.96
	TEFF	TSE	Indir	0.004	0.96
$r^2$	.95				
Within	TEFF	CM	Dir	0.65	<0.001***
		TSE	Dir	-0.01	0.64
	CM	TSE	Dir	0.01	0.79
	TEFF	TSE	Indir	0.01	0.79
$r^2$	.43				

*Note.*  $p > .05$  nonsignificant; \* is  $p \leq .05$ ; \*\* is  $p \leq .01$ ; \*\*\* is  $p \leq .001$

Figure 2. Model #1: Direct and Indirect Effects



Note: Indirect effects are reported in parentheses

Table 2 and Figure 2 represent direct and indirect effects in Model 1. The only significant direct effect was between observer-reported classroom management and observer-reported TE. Observer-reported classroom management predicted observer-reported TE at the between- and within-teacher level. Thus, teachers who were rated by observers to be good managers tended to also be rated by observers as having higher TE,  $\beta = .97, p < .001$ . Within teachers, days when teachers were rated as better at managing their classroom tended to be days that they were rated by observers to be more effective,  $\beta = .65, p < .001$ . No significant direct effects were found between TSE and observer-reported teaching efficacy at the between or within level. Teachers'

self-ratings of their teaching efficacy was not predictive of observer ratings of classroom management nor observer ratings of their TE,  $\beta = .004, p = .96$ . Furthermore, within-teachers, teachers TSE scores were not predictive of days they were rated by observers to be more effective classroom managers nor more effective at teaching,  $\beta = .01, p = .79$ . Finally, no significant indirect relationship was found between TSE and observer-reported TE at the within- or between-teacher level ( $\beta = .004, p < .96; \beta = .01, p < .79$ ).

**Model #2: Observer-reported TE and teacher self-reported classroom management.**

Table 3 presents the means and interclass correlation coefficients (ICC) as well as the between- and within-teacher correlations and standard deviations of the second model.

Table 3  
*Model #2 Correlations, Means, Standard Deviations, and ICC*

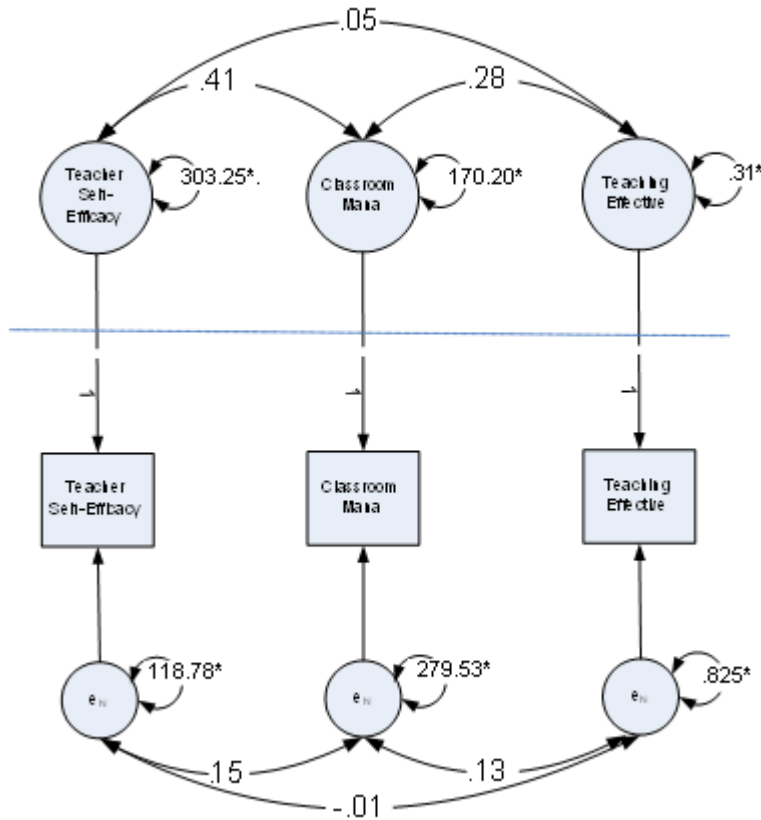
	TSE	CM	TEFF
TSE	1	.41*	.05
CM	.15*	1	.28*
TEFF	-.01	.13*	1
Mean	127.81	79.62	3.60
SD between	17.41	13.05	0.56
SD within	10.90	16.72	0.91
ICC	.72	.38	.27

The mean score among teachers on the NTSES, was 127.81, indicating that, on average, teachers tended to rate themselves as highly efficacious in teaching. An ICC of .72 also suggests that teachers' TSE ratings tended to be stable within teachers. The mean score of observer-reported classroom management among teachers was 79.62 out of a total of 100 possible points on a visual analog scale. Teachers tended to rate their ability to manage a classroom as high. The average rating for observer-reported TE was 3.60, which is slightly above the mid-point of the scale. Self-reported classroom management ICC was .38 and, similar to Model 1, observer-



reported teaching effectiveness was .27. Both ICCs indicate that observer ratings tended to be less stable between teachers with the majority of the variability being within teachers.

*Figure 3. Model #2 Variable Correlations*



A Pearson correlations coefficient was computed to assess the relations between the variables in Model 2. Figure 3 represents the correlations between variables, as well as variance among the different variables. Between-teachers, there was a significant moderate positive correlation between TSE and self-reported classroom management indicating that teachers who rated themselves as highly efficacious at their jobs also tended to rate themselves highly in classroom management,  $r = .41, p < .001$ . However, at the within-teacher level, the correlation,

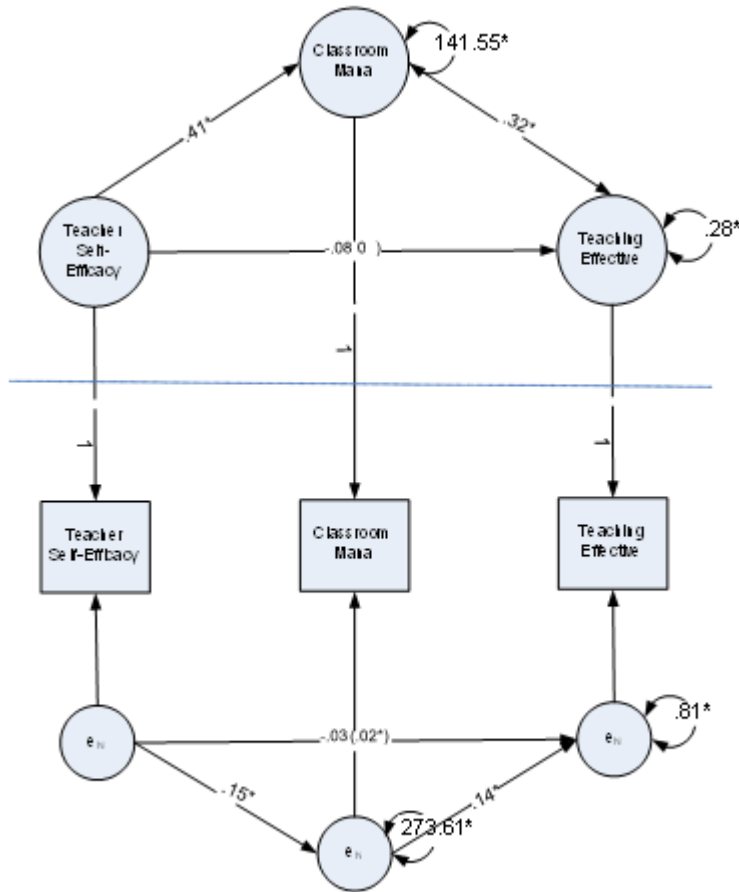
while still significant, was weaker,  $r = .15$ ,  $p < .001$ . Thus at the within-teacher level, a teacher's TSE was less correlated with their multiple ratings of their own classroom management on different days. A significant positive relationship was found at the between-teacher level between self-reported classroom management and observer-reported TE ( $r_{\text{between}} = .28$ ,  $p < .01$ ) and a significant but weaker positive correlation was found at the within-teacher level ( $r_{\text{within}} = .13$ ,  $p < .001$ ). At the between-teacher level a nonsignificant correlation ( $r_{\text{between}} = .05$ ,  $p = .66$ ) was found for TSE and observer-reported TE, and at the within-teacher level the correlation was also nonsignificant ( $r_{\text{within}} = -.01$ ,  $p = .88$ ).

Table 4  
*Model #2: Direct and Indirect Effects*

	<b>Outcome</b>	<b>Predictor</b>	<b>Relationship</b>	<b>Est.</b>	<b>P</b>
Between	TEFF	CM	Dir	.32	<0.001***
		TSE	Dir	-.08	.43
	CM	TSE	Dir	.41	0.001***
	TEFF	TSE	Indir	.13	.01**
$r^2$	.09				
Within	TEFF	CM	Dir	.14	<0.001***
		TSE	Dir	-.03	.46
	CM	TSE	Dir	.15	<0.001***
	TEFF	TSE	Indir	.02	.01**
$r^2$	.02				

Note.  $p > .05$  nonsignificant; \* is  $p \leq .05$ ; \*\* is  $p \leq .01$ ; \*\*\* is  $p \leq .001$

Figure 4. Model #2: Direct and Indirect Effects



Note: Indirect effects are reported in parentheses

Table 4 and Figure 4 represent direct and indirect effects of Model 2. Three significant direct effects and one indirect effect were found in this model at both the between- and within-teacher level. There was a significant direct relationship between self-reported classroom management and observer-reported TE at the between- and within-teacher levels. Teachers who rated themselves as good classroom managers were more likely to be rated by observers as having high TE,  $\beta = .32, p < .001$ . Within teachers, on days when teachers rated themselves as better at managing their classrooms, also tended to be days when observers rated them as having higher TE,  $\beta = .14, p < .001$ . There was also a significant direct relationship between TSE and

self-reported classroom management. Teachers who rated themselves as having high TSE typically also rated themselves as having higher self-reported classroom management,  $\beta = .41, p < .001$ . Furthermore, within-teachers, high TSE scores predicted a greater frequency of days where teachers rated themselves as high on classroom management skills,  $\beta = .15, p < .001$ . No significant direct effect was found at the between- or within-teacher level for the relation between TSE and observer-reported TE ( $\beta_{\text{between}} = -.08, p = .43; \beta_{\text{within}} = -.03, p = .46$ ). However, a significant indirect relation was found between TSE and observer-reported teaching effectiveness. Teachers who rated themselves higher in TSE tended to indirectly predict higher observer-reported scores on TE via classroom management,  $\beta = .13, p = .01$ . In other words, although TSE did not directly impact TE, TSE did influence classroom management and this influence on classroom management impacted TE. Likewise, within-teachers, TSE was found to indirectly predict observer ratings of TE, via teacher reported classroom management,  $\beta = .02, p = .01$ .

**Model #3: Self-reported TE and observer-reported classroom management model.**

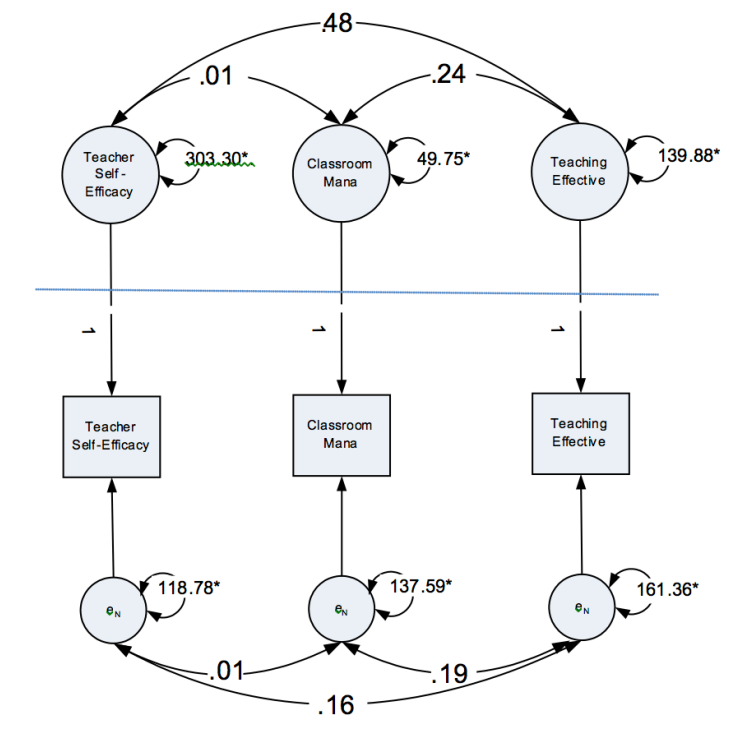
Table 5 presents the means and interclass correlation coefficients (ICC) as well as the between- and within-teacher correlations and standard deviations of the third model.

Table 5  
*Model #3 Correlations, Means, Standard Deviations, and ICC*

	TSE	CM	TEFF
TSE	1	.01	.48*
CM	.01	1	.24*
TEFF	.16*	.19*	1
Mean	127.81	70.58	80.71
SD between	17.42	7.05	11.83
SD within	10.90	11.73	12.70
ICC	.72	.27	.46

Similar to Model 2 the mean score among teachers on the NTSES was 127.81, indicating that on average, teachers tended to rate themselves highly in regards to TSE. Also, similar to previous models, observer-reported classroom management had a mean of 70.58 indicating that teachers generally tended to rate their ability to manage a classroom as high. Self-reported TE was 80.71 out of a possible 100, which suggests that, on average, teachers rate themselves high on effectiveness. ICCs of TSE and observer-reported classroom management were also similar to prior models, however, self-reported teaching efficacy was .46 indicating that it is slightly more stable between- teachers when compared to observer-reported TE (ICC = .27).

Figure 5. Model #3 Variable Correlations



A Pearson correlation coefficient was computed to assess the relations between the variables in Model 3. Figure 5 represents the correlations between variables, as well as the variance among different variables. A significant and moderate positive correlation was found between TSE and self-reported TE at the between-teacher level,  $r = .48$ ,  $p < .001$ . At the within-

teacher level the significant positive correlation was weaker,  $r = .16, p < .001$ . Additionally, at the between-teacher level there was a significant positive correlation between observer-reported classroom management ratings and observer-reported TE ( $r_{\text{between}} = .24, p = .02$ ) and a significant within-teacher level positive correlation ( $r_{\text{within}} = .19, p < .001$ ). Finally, between- and within-teacher level correlations between TSE and observer-reported classroom management, were nonsignificant ( $r_{\text{between}} = .01, p = .99$ ;  $r_{\text{within}} = .01, p = .71$ ).

Table 6

*Model #3: Direct and Indirect Effects*

		<b>Relationships</b>			
	<b>Outcome</b>	<b>Predictor</b>	<b>p</b>	<b>Est.</b>	<b>P</b>
Between	TEFF	CM	Dir	.23	.01**
		TSE	Dir	.48	<0.001***
	CM	TSE	Dir	.01	.92
	TEFF	TSE	Indir	.002	.92
$r^2$	.28				
Within	TEFF	CM	Dir	.19	<0.001***
		TSE	Dir	.16	<0.001***
	CM	TSE	Dir	.01	.75
	TEFF	TSE	Indir	.002	.75
$r^2$	.06				

*Note.*  $p > .05$  nonsignificant; \* is  $p \leq .05$ ; \*\* is  $p \leq .01$ ; \*\*\* is  $p \leq .001$

Figure 6. Model #3: Direct and Indirect Effects

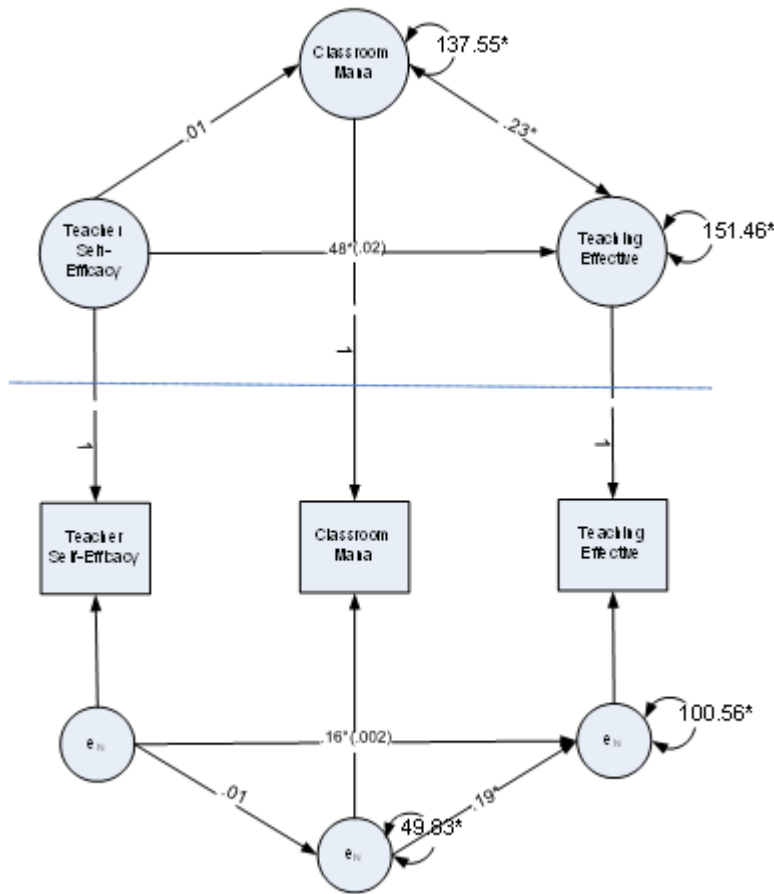


Table 6 and Figure 6 present the direct and indirect relations for Model 3. Two significant direct relations were found between variables at the between- and within-teacher level. High observer ratings of classroom management tended to predict high self-ratings of TE at the between-teacher level,  $\beta = .23, p = .01$ . At the within-teacher level, days when observers rated individual teachers high on classroom management tended to be days that teachers reported higher levels of their own TE,  $\beta = .19, p < .001$ . Additionally, at the between-level, the teachers who rated themselves high in TSE were the same teachers who rated themselves highly in TE,  $\beta = .48, p < .001$ . Similarly, at the within-teacher level, when teachers rated themselves high in TSE they also tended to rate themselves highly in TE. Conversely, there was no significant direct

relation among TSE and observer-reported classroom management at the between- or within-teacher level ( $\beta_{\text{between}} = .01, p = .92$ ;  $\beta_{\text{within}} = .01, p = .75$ ). Finally, no indirect relation was found between TSE and self-reported TE at the between- or within-teacher level ( $\beta_{\text{between}} = .02, p = .92$ ;  $\beta_{\text{within}} = .002, p = .75$ ).

**Model 4: Self-reported TE and classroom management model.** Table 7 presents the means and interclass correlation coefficients (ICC), as well as the between- and within-teacher correlations and standard deviations of the fourth model.

Table 7  
*Model #4 Correlations, Means, Standard Deviations, and ICC*

	TSE	CM	TEFF
TSE	1	.42*	.48*
CM	.14*	1	.94*
TEFF	.16*	.78*	1
Mean	127.84	79.64	80.70
SD between	17.42	13.04	11.84
SD within	10.89	16.72	12.70
ICC	.72	.38	.47

Similar to prior models, teachers generally rated themselves high in TSE ( $M = 127.84$ ) and classroom management ( $M = 79.64$ ). Regarding self-reported TE was 80.70 out of a possible 100, which suggests that, on average, teachers rate themselves high on effectiveness. Similar to prior models the ICC for TSE was .72. Self-reported classroom management was .38, which tended to be more stable than its observer-reported counterpart ( $ICC = .27$ ) seen in other models. Similarly, self-reported TE tended to have higher ICCs than observer-reported TE ( $ICC_{\text{self}} = .47$ ;  $ICC_{\text{observer}} = .27$ ).

Figure 7. Model 4 Variable Correlations



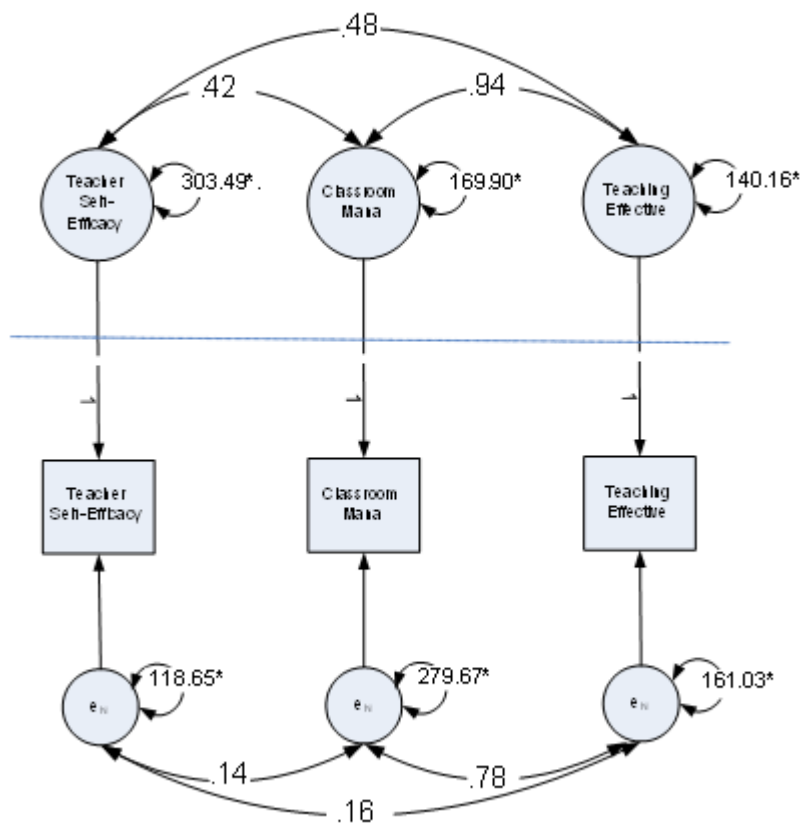


Figure 7 shows the correlations between the different variables in Model 4. Compared to previous models, this fourth model has many more moderate to strong correlations than previous ones. In fact, all correlations at both the within and between levels are significant in Model 4. This indicates that, at least, when rated by the same rater, these variables tend to be correlated. At the between-teacher level a very strong positive correlation was found between self-reported classroom management and self-reported TE,  $r = .94, p < .001$ . At the within-teacher level this positive correlation was strong,  $r = .78, p < .001$ . In general, teachers who rated themselves highly in classroom management skills also rated their TE to be high. Furthermore, days when teachers rated themselves as being high in classroom management were also days when they rated themselves highly in effective teaching. There was a moderate positive correlation at the

between-teacher level and a weaker within-teacher level correlation between TSE and self-ratings of classroom management ( $r_{\text{between}} = .42, p < .001; r_{\text{within}} = .14, p < .001$ ). While, in general, teachers' TSE scores were positively correlated with their classroom management ratings, TSE had less influence on a teacher's classroom management ratings on different days. Similarly, a moderate positive correlation was found at the between-teacher level and a weaker positive correlation was found at the within-teacher level between TSE and self-reported TE ( $r_{\text{between}} = .48, p < .001; r_{\text{within}} = .16, p < .001$ ).

Table 8  
*Model #4: Direct and Indirect Effects*

	<b>Outcome</b>	<b>Predictor</b>	<b>Relationship</b>	<b>Est.</b>	<b>P</b>
Between	TEFF	CM	Dir	.89	<0.001***
		TSE	Dir	.11	.01**
	CM	TSE	Dir	.42	<0.001***
	TEFF	TSE	Indir	.37	<0.001***
$r^2$	.89				
Within	TEFF	CM	Dir	.78	<0.001***
		TSE	Dir	.05	0.03*
	CM	TSE	Dir	.14	<0.001***
	TEFF	TSE	Indir	.11	<0.001***
$r^2$	.61				

*Note.*  $p > .05$  nonsignificant; \* is  $p \leq .05$ ; \*\* is  $p \leq .01$ ; \*\*\* is  $p \leq .001$

Figure 8. Model #3: Direct and Indirect Effects

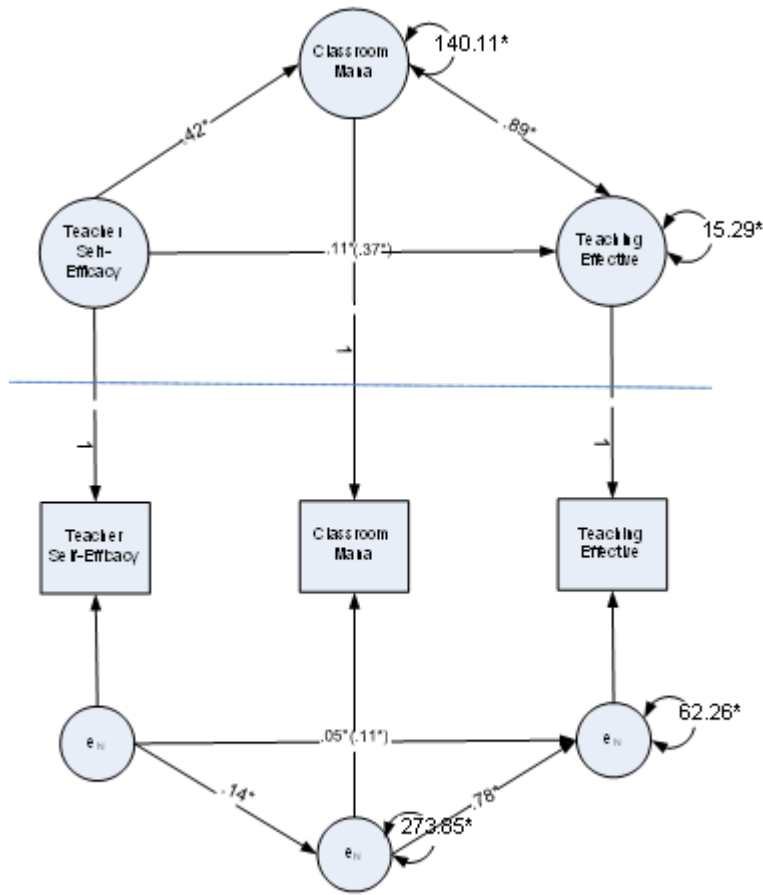


Table 8 and Figure 8 show the direct and indirect relations among variables in Model 4. There were three significant direct effects within this model. Self-reported classroom management significantly predicted self-reported TE at the between- and within-teacher levels ( $\beta_{\text{between}} = .89, p < .001; \beta_{\text{within}} = .78, p < .001$ ). When teachers rated themselves highly in classroom management they also tended to rate themselves high in TE. Furthermore, within teachers, days when teachers rated themselves high in classroom management also tended to be days when they rated themselves as being more effective teachers. Both at the between and within-teacher level, teachers high in TSE tended to also rate themselves high in classroom management ( $\beta_{\text{between}} = .42, p < .001; \beta_{\text{within}} = .14, p < .001$ ). High TSE significantly predicted

self-reported classroom management among the sample of teachers. Furthermore, high TSE predicted higher individual day-to-day ratings on self-reported classroom management. Additionally, there a significant direct effect was found at the between- and within-teacher level among TSE and self-reported TE ( $\beta_{\text{between}} = .11, p = .01; \beta_{\text{within}} = .05, p = .03$ ). At the between-teacher level, higher TSE scores tended to predict higher self-reported TE scores. At the within-teacher level, higher TSE scores predicted higher day-to-day ratings of individual teachers TE. Finally, a significant indirect effect was found between TSE and self-reported TE at both the within- and between-teacher level ( $\beta_{\text{between}} = .37, p < .001; \beta_{\text{within}} = .11, p < .001$ ).

## **Hypotheses**

This section will summarize the results in terms of the hypotheses.

**Hypothesis 1: TSE and TE.** My first hypothesis focused on how well TSE predicted teaching effectiveness (TE). Specifically, I predicted that TSE will directly predict both self-reported TE and observer-reported TE. Additionally, I examined the relation at both the between and within-teacher level. As seen in the tables and figures above, this hypothesis received mixed support. Specifically, Models 1 (observer-reported TE and classroom management) TSE was not able to directly predict observer-reported TE at either the between- or within-teacher level. However, Models 3 (self-reported TE and observer-reported classroom management) and 4 (self-reported TE and classroom management), which used teacher self-report of TSE, did directly predict self-reported TE at both the between- and within-teacher level. These findings suggest that TSE only predicts TE when TE is measured through self-reports.

**Hypothesis 2: TSE and classroom management.** My second hypothesis focused on how well TSE predicted classroom management. Specifically, I predicted that TSE would directly predict both self- and observer reported classroom management. Much like the first

hypothesis, TSE only directly predicted self-reported classroom management at the between- and within-teacher level. Model 1(observer-reported TE and classroom management) and Model 3 (self-reported TE and observer-reported classroom management), which both use observed classroom management, show a nonsignificant direct relationship between TSE and observer-reported classroom management at both the between- and within-teacher level. These findings suggest that TSE only predicts classroom management when classroom management is measured through self-reported means.

**Hypothesis 3: TSE indirect impact on TE via classroom management.** My final hypothesis focused on the indirect impact of TSE on TE. Specifically, I predicted that TSE will have an indirect impact on both self- and observer-reported TE via both self- and observer-reported classroom management. This hypothesis also revealed mixed results. An indirect relationship of TSE on TE was found at the between- and within-teacher level for Model 2 (observer-reported TE and self-reported classroom management) and 4 (self-reported TE and classroom management), which measures TE using observer report. TSE indirectly impacted observer-reported TE via self-reported classroom management in Model 2. In Model 4 TSE indirectly impacted self-reported TE via self-reported classroom management at both the between- and within-teacher level. These results likely suggest that there is an indirect impact of TSE on TE only when classroom management is measured via teacher self-reports.

As stated previously in this paper, given the conservative nature of the Sobel test, nonsignificant indirect effects were to be followed up with the Baron and Kenny (1986) method. Nonsignificant indirect effects were found for both Model 1 (observer-reported TE and classroom management) and Model 3 (self-reported TE and observer-reported classroom management). The Baron and Kenny (1986) method was not utilized for Model 1 given that

TSE was not significantly related to either observer-reported teaching effectiveness or observer-reported classroom management at the between- or within-teacher level, both steps of this method were not met and thus the test could not be performed. Similarly, in Model 3 TSE did not significantly predict observer-reported classroom management at the between- or within-teacher level, which does not meet step two of the Baron and Kenny (1986) method and thus nonsignificant results could not be investigated using this Baron and Kenny method.

### **Discussion**

It is widely acknowledged that teacher quality directly relates to student academic outcomes (Darling-Hammond, 2000; Goe, 2007). Having an effective teacher can result in substantial academic gains over the course of a school year. Yet, it is still not clear why some teachers are able to promote student success while others do not (Darling-Hammond, 2000; Hanushek & Rivkin, 2010). An emerging body of research has found mixed, but promising results for the relationship between TSE and teacher quality. The literature to-date has noted relations between TSE in the research and certain effective teaching practices such as classroom management, that are known to significantly relate to quality teaching. However, most notable is the fact that TSE has failed to yield significant correlations with teacher quality overall. This may be due to the methodological limitations of these studies that are often fraught with small sample sizes, lack of longitudinal studies, and retrospective self-reports of instructional effectiveness. Additionally, few studies to-date have explored the potential mediating variables that impact the relationship between TSE and teacher quality. To begin to address these gaps in the literature, the proposed study sought to determine if TSE has an indirect impact on teaching effectiveness (TE) via classroom management.

Descriptive analyses revealed that in general, teachers rated themselves as being high in TSE. Given that the average years spent teaching for this population was 12, this finding is not surprising since, in general, more experienced teachers tend to have higher TSE (Tschannen-Moran & Hoy, 2007). Consistent with the notion that these were experienced and competent teachers, the average score for observer- and self-reported classroom management and TE also tended to be high. Correlational models indicated that both self- and observer-reported classroom management was significantly related to TE. The strength of this relationship was strongest in Model 1 (observer-reported TE and classroom management) and Model 4 (self-reported TE and classroom management) both of which had the same rater for classroom management and TE. Thus, the present study indicated that when raters are consistent, there is a significant strong positive correlation between classroom management and TE at both the between and within-teacher level. This is consistent with the notion of shared method variance from using mono-operations, and justifies the need to have multiple reporters to avoid bias in estimating relationships.

In Model 1, the between level correlation was .97 between observer-reported classroom management and observer-reported TE indicating that, in this model, classroom management and TE were approximately interchangeable. In other words, observers' ratings of teachers' classroom management practices were nearly identical to observers' ratings of that teacher's TE. Notably, significant, moderate positive correlations between-teachers occurred between TSE and classroom management only when classroom management was self-reported.

The three hypotheses of the present study were investigated via the direct and indirect effects for each of the four models. Regarding the first hypothesis, between teachers, TSE scores predicted TE scores when TSE was self-reported. The between-teacher findings were also

consistent at the within-teacher level whereby high TSE scores tended to predict more consistent high scores day-to-day on self-reported classroom management (both observed and self-reported).

TSE scores did not predict TE at the between- or within-teacher level when TE was rated by observers. These findings are consistent with the present body of research on TSE where the significance of relations between TSE and TE have generally been mixed (Allinder, 1995; Guo, Piasta, Justice, Kadaverek, 2010; Martin, Sass, & Schmitt, 2012; Pakarinen et al., 2010; Thoonen, Slegers, Oort, Peetsma, & Geijsel, 2011; Weshah, 2012).

The second hypothesis regarding TSE directly predicting self- and observer-reported classroom management also received mixed support. At the between- and within-person level, TSE was found to only predict classroom management as reported by teachers. In general, teachers with high TSE scores tended to also have higher self-reported classroom management ratings. TSE did not significantly predict observer reports of teacher's classroom management practices at the between- or within-teacher level. This contrasts with a handful of studies to-date that have found significant direct effects of TSE on observer reported classroom management (Morris-Rothschild, 2006; Abu-Tineh, Khasawneh, & Khalaileh, 2011). One reason for this may be due to the high correlation of scores (.97) between observer reports of classroom management and observer reports of TE that was found in Model 1. It is possible that observers' ratings of individual teacher's TE were much more aligned with classroom management in this study when compared to others. Given that the observational instrument seeks to document multiple aspects of TE within a single observation, its use may have been burdensome to raters. This could in turn have made observers rely on more easily apparent aspects of TE such as classroom management rather than the other three domains measured by the instrument.



The third and final hypothesis, concerning the indirect relationship between TSE and TE, was partially supported. Two models, where classroom management was self-reported, provide support for the indirect effects of TSE on TE. Self-reported classroom management was found to fully mediate the relationship between TSE and observer-reported TE. When teachers report high TSE and high classroom management skills, observers also tended to rate these teachers as being highly effective in teaching. This finding was consistent at the between- and within-teacher level. Furthermore, self-reported classroom management was found to partially mediate the relationship between TSE and self-reported TE. However, observer-reported classroom management did not mediate the relationship between TSE and self- and observer reported TE at the between- and within-teacher level.

Aside from the study's three hypotheses, it should be noted that in each of the four models, classroom management, whether self- or observer-reported, always significantly predicted self- and observer-ratings of teaching effectiveness. As previously mentioned, significant correlations were found between these two variables for each of the four models. Classroom management is significantly correlated and is a good predictor of effective teaching, whether it be teacher or observer rated. While this finding is supported by an established body of literature, it is worth mentioning especially given the uniqueness and strength of the study design.

### **Strengths and Limitations**

While there are many strengths of this study there are several limitations that should be considered in light of the present results. One limitation is that TSE is only measured four times during the original data collection procedures while both observer and self-reported classroom management were measured up to eight time points. Additionally, both classroom observations

of teachers and teachers' momentary self-reports occurred during three-day periods during each wave of data collection whereas the self-reported measure of TSE did not occur during this time, but typically prior. Thus, there is a possibility that self-reported TSE scores may not be as precise as if they were measured at the same time as observer and self-ratings of classroom management and TE and as a result were unable to show significant direct relations to these variables. Bandura (1997), conceptualized self-efficacy as an innate construct within an individual that builds, adapts, and changes based on four environmental sources (e.g. mastery experiences, vicarious experiences, verbal persuasion, and physiological feedback). Thus, the teacher's TSE during these periods of data collection could have increased or decreased due to one or more of these environmental sources. However, it should be noted that among veteran teachers, TSE has been found to be relatively stable (e.g. more than five years teaching experience) and thus not as malleable compared to novice teachers (Bandura, 1997; Tschannen-Moran & Hoy, 2007). Nevertheless, there was clearly method variance in the data that did not correspond across sources and it is possible that the constructs were also variable across measurement circumstances.

A second limitation of this study, and many other studies on TSE, is that TSE was measured as a global construct. Bandura (1997) cautions against measuring SE as global because SE is task-specific. He also argued that by measuring TSE from a multifaceted perspective one would increase the measures predictive validity. If the present study had looked at the specific TSE domains within these models it is likely that some of the domains may have fit within these models better than others and the global measure of TSE. This investigation was beyond the scope of the original study, and can be undertaken as a follow-up study. Thus, these findings are

all the more meaningful, as the current design probably underestimates the role of self-efficacy in predicting or explaining teaching effectiveness.

### **Implications**

The present study partially supports the notion that TSE indirectly predicts TE via classroom management. Given that there is support for both the direct and indirect effects of TSE on TE and direct effects between TSE and classroom management, TSE is a promising variable for intervention. When teachers fail to meet achievement benchmarks and receive low supervisor ratings, one means of intervention may be to build upon skill acquisition while simultaneously monitoring and promoting their self-efficacy in teaching. Additionally, the results of this study can guide school psychologists in increasing the effectiveness of their consultative services, especially in the area of classroom management, provided to teachers.

Consultation within the schools often takes place between the school psychologist and the teachers to address a variety of concerns, most prominently those related to classroom management and organization (MacSuga & Simonsen, 2011; Motoca et al., 2014). However, consultation by itself does not directly translate to successful implementation by the teacher. The literature on consultation points to several factors that increase the likelihood of interventions success of implementation, one of them being improving self-efficacy (Allinder, 1994; Capizzi, Wehby, & Sandmel, 2010; Hagen et al., 1998). Thus, consultants should take into consideration teachers' TSE in order to determine the probability of successful implementation. Furthermore, school psychologists may consider methods to increase teacher self-efficacy during consultation such as vicarious experience and verbal persuasion. By addressing TSE during consultation, school psychologists can potentially increase the probability of teachers implementing interventions with fidelity, consistency, and persistence, which may lead to better student outcomes.

Importantly, the results of this study provide further support for the inexorably tangled relationship of classroom management and TE. Thus, it provides further support of the importance and

necessity of districts and teacher training institutions to divert more resources into training preservice and inservice teachers' in classroom management. The literature consistently shows that classroom management is not emphasized or is inadequate in teacher training programs (Emmer & Stough, 2001; Stough, 2006). Furthermore, many teachers report little opportunity for continuing professional development in this area. The findings from this study support the notion that better classroom managers tend to also be better teachers, thus it is essential that future educators be equipped with skills to effectively manage their classrooms.

### **Future Directions**

Future studies using this sample and other multifaceted measures of TSE will want to explore how different, task specific domains of TSE can impact the direct and indirect relations within these models. In particular, the TSE domain of classroom management should be investigated. It is very likely that classroom management teaching self-efficacy would have stronger direct and indirect effects than were found in the current study, which used general TSE to predict CM and TE. Bandura (1997) cautions against omnibus measures of TSE as they may be less predictive than more specific TSE domains. Thus, domain specific TSE may provide more support for these mediation models than overall TSE. This should also be considered in teacher training, as training to specific tasks will be more impactful on TSE for that task than TSE in general. Thus, specific tasks should get specific attention with distinct, task-specific activities, measures, and discussion using four task-specific sources of influence on TSE: direct experience, vicarious learning, discussion and social persuasion, and affect management.

## References

- Abu-Tineh, A. M., Khasawneh, S. A., & Khalailah, H. A. (2011). Teacher self-efficacy and classroom management styles in Jordanian schools. *Management in Education, 25*(4), 175-181.
- Ahn, S., & Choi, J. (2004). Teachers' Subject Matter Knowledge as a Teacher Qualification: A Synthesis of the Quantitative Literature on Students' Mathematics Achievement. *Online Submission*.
- Aili, C., & Brante, G. (2007). Qualifying teacher work: everyday work as basis for the autonomy of the teaching profession. *Teachers and Teaching: theory and practice, 13*(3), 287-306.
- Allinder, R. M. (1994). The relationship between efficacy and the instructional practices of special education teachers and consultants. *Teacher Education and Special Education, 17*(2), 86-95.
- Andrews, J. W. (1980). Preservice Performance and the National Teacher Exams. *Phi Delta Kappan, 61*(5), 358-359.
- Armor, D., Conroy-Oseguera, P., Cox, M., King, N., McDonnell, L., & Pascal, A. P. E., & Zellman, G. (1976). *Analysis of the school preferred reading programs in selected Los Angeles minority schools*.
- Artelt, C., & Schneider, W. (2015). Cross-Country Generalizability of the Role of Metacognitive Knowledge in Students' Strategy Use and Reading Competence. *Teachers College Record, 117*(1), n1.
- Ashton, P., & Webb, R. (1982). *Teachers' sense of efficacy: Toward an ecological model*. Paper presented at the annual meeting of the American Educational Research Association, New York.

- Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement*: Longman Publishing Group.
- 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*, 69-78.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*: Macmillan.
- Baumert, J., & Kunter, M. (2013). The COACTIV Model of Teachers' Professional Competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers: Results from the COACTIV Project* (pp. 25-48). Boston, MA: Springer US.
- Beatty-O'Ferrall, M. E., Green, A., & Hanna, F. (2010). Classroom management strategies for difficult students: Promoting change through relationships. *Middle School Journal, 41*(4), 4-11.
- Becker, W. C., Madsen Jr, C. H., Arnold, C. R., & Thomas, D. R. (1967). The contingent use of teacher attention and praise in reducing classroom behavior problems. *The Journal of Special Education, 1*(3), 287-307.
- Benight, C. C., & Bandura, A. (2004). Social cognitive theory of posttraumatic recovery: The role of perceived self-efficacy. *Behaviour research and therapy, 42*(10), 1129-1148.

- Berman, P. S., MacLaughlin, M. W., & Bass, G. (1977). *Factors affecting implementation and continuation*: Rand.
- Berry, R. A. (2010). Preservice and Early Career Teachers' Attitudes toward Inclusion, Instructional Accommodations, and Fairness: Three Profiles. *Teacher Educator, 45*(2), 75-95.
- Betts, J. R., Zau, A., & Rice, L. (2003). *Determinants of student achievement: New evidence from San Diego*: Public Policy Institute of California San Francisco.
- Black, D. W. (2017). Abandoning the Federal Role in Education: The Every Student Succeeds Act. *Calif. L. Rev., 105*, 1309.
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary Methods: Capturing Life as it is Lived. *Annual Review of Psychology, 54*, 579-616.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. In: MIT Press.
- Brophy, J. (1988). Educating teachers about managing classrooms and students. *Teaching and Teacher Education, 4*(1), 1-18.
- Brophy, J., Good, T., & Wittrock, M. (1986). Handbook of research on teaching. In (pp. 328-375).
- Cantrell, P., Young, S., & Moore, A. (2003). Factors affecting science teaching efficacy of preservice elementary teachers. *Journal of Science Teacher Education, 14*(3), 177-192.
- Capizzi, A. M., Wehby, J. H., & Sandmel, K. N. (2010). Enhancing mentoring of teacher candidates through consultative feedback and self-evaluation of instructional delivery. *Teacher Education and Special Education, 33*(3), 191-212.

- 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.
- Carr, D. (2013). The Effects of Teacher Preparation Programs on Novice Teachers Regarding Classroom Management, Academic Preparation, Time Management, and Self Efficacy.
- Carson, R. L., Weiss, H. M., & Templin, T. J. (2010). Ecological momentary assessment: A research method for studying the daily lives of teachers. *International Journal of Research & Method in Education, 33*(2), 165-182.
- Cason, M. F. (2018). *The Impact of Student Engagement, Instructional Strategies, and Classroom Management on Self-Efficacy of Christian Private School Teachers.*
- Cawelti, G. (2004). Handbook of research on improving student achievement. *Educational Research Service.*
- Chacon, C. T. (2005). Teachers' perceived efficacy among English as a foreign language teachers in middle schools in Venezuela. *Teaching and Teacher Education, 21*(3), 257-272.
- Chang, T. S., Lin, H. H., & Song, M. M. (2011). University faculty members' perceptions of their teaching efficacy. *Innovations in Education and Teaching International, 48*(1), 49-60.
- Chong, W. H., Klassen, R. M., Huan, V. S., Wong, I., & Kates, A. D. (2010). The relationships among school types, teacher efficacy beliefs, and academic climate: Perspective from Asian middle schools. *The Journal of Educational Research, 103*(3), 183-190.
- Christenson, S. L., Reschly, A. L., & Wylie, C. (2012). *Handbook of research on student engagement*: Springer Science & Business Media.



- Clare, L., Valdés, R., Pascal, J., & Steinberg, J. R. (2001). Teachers' assignments as indicators of instructional quality in elementary schools. *Center for the Study of Evaluation Technical Report, 545*.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects. *Journal of Human Resources, 45*(3), 655-681.
- Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25*(2), 119-142.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences, 2nd Edition*. Hillsdale, N.J: Lawrence Erlbaum.
- Coladarci, T. (1992). Teachers' sense of efficacy and commitment to teaching. *The Journal of Experimental Education, 60*(4), 323-337.
- Coladarci, T., & Breton, W. A. (1997). Teacher efficacy, supervision, and the special education resource-room teacher. *The Journal of Educational Research, 90*(4), 230-239.
- Colvin, G., Kameenui, E. J., & Sugai, G. (1993). Reconceptualizing behavior management and school-wide discipline in general education. *Education and treatment of children, 361-381*.
- Conner, A., Singletary, L. M., Smith, R. C., Wagner, P. A., & Francisco, R. T. (2014). Teacher support for collective argumentation: A framework for examining how teachers support students' engagement in mathematical activities. *Educational Studies in Mathematics, 86*(3), 401-429.
- Corbett, D., & Wilson, B. (2002). What urban students say about good teaching. *Educational Leadership, 60*(1), 18-23.

- Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research, 77*(1), 113-143.
- Cornell, D. G., & Mayer, M. J. (2010). Why do school order and safety matter? *Educational researcher, 39*(1), 7-15.
- Cotton, S. (2002). The training needs of vocational teachers for working with learners with special needs.
- Covino, E. A., & Iwanicki, E. F. (1996). Experienced teachers: Their constructs of effective teaching. *Journal of Personnel Evaluation in Education, 10*(4), 325-363.
- Daniels, K., Hartley, R., & Travers, C. J. (2006). Beliefs about stressors alter stressors' impact: Evidence from two experience-sampling studies. *Human Relations, 59*(9), 1261-1285.
- Darling-Hammond, L. (2000). Teacher quality and student achievement. *education policy analysis archives, 8*, 1.
- Darling-Hammond, L. (2003). Keeping good teachers: Why it matters, what leaders can do. *Educational Leadership, 60*(8), 6-13.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan, 93*(6), 8-15.
- Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *education policy analysis archives, 13*, 42.
- Darling-Hammond, L., & Sykes, G. (2003). Wanted, a national teacher supply policy for education: The right way to meet the "highly qualified teacher" challenge. *education policy analysis archives, 11*, 33.

- Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics*, 86(1), 195-210.
- Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources*, 42(3), 528-554.
- DeJarnette, N. K., & Sudeck, M. (2015). Supporting clinical practice candidates in learning community development. *Teacher Development*, 19(3), 311-327.
- Dembo, M. H., & Gibson, S. (1985). Teachers' sense of efficacy: An important factor in school improvement. *The Elementary School Journal*, 86(2), 173-184.
- Demmon-Berger, D. (1986). *Effective Teaching: Observations from Research*: ERIC.
- Desimone, L. M., Smith, T. M., & Frisvold, D. E. (2010). Survey measures of classroom instruction: Comparing student and teacher reports. *Educational Policy*, 24(2), 267-329.
- Dicke, T., Parker, P. D., Marsh, H. W., Kunter, M., Schmeck, A., & Leutner, D. (2014). Self-efficacy in classroom management, classroom disturbances, and emotional exhaustion: A moderated mediation analysis of teacher candidates. *Journal of Educational Psychology*, 106(2), 569.
- Diener, E., Smith, H., & Fujita, F. (1995). The personality structure of affect. *Journal of personality and social psychology*, 69(1), 130.
- Djigic, G., & Stojiljkovic, S. (2011). Classroom management styles, classroom climate and school achievement. *Procedia-Social and Behavioral Sciences*, 29, 819-828.
- Donovan, M., Bransford, J., & Pellegrino, J. (1999). How people learn: Bridging research and practice, committee on learning, research and educational practice. *Washington DC: National Research Council*.

- Doyle, W. (1986). Content representation in teachers' definitions of academic work. *Journal of Curriculum Studies, 18*(4), 365-379.
- Dukerich, L. (2015). Applying modeling instruction to high school chemistry to improve students' conceptual understanding. *Journal of Chemical Education, 92*(8), 1315-1319.
- Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation, 14*(7), 1-11.
- Egalite, A. J., Kisida, B., & Winters, M. A. (2015). Representation in the classroom: The effect of own-race teachers on student achievement. *Economics of Education review, 45*, 44-52.
- Ehrenberg, R. G., Goldhaber, D. D., & Brewer, D. J. (1995). Do teachers' race, gender, and ethnicity matter? Evidence from the National Educational Longitudinal Study of 1988. *ILR Review, 48*(3), 547-561.
- Emmer, E., Evertson, C., Clements, B., & Worsham, M. (1997). Managing problem behaviors. *Classroom Management for Secondary Teachers*.
- Emmer, E., & Sabornie, E. (2013). *Handbook of classroom management: Research, practice, and contemporary issues*: Routledge.
- Emmer, E. T., & Hickman, J. (1991). Teacher efficacy in classroom management and discipline. *Educational and Psychological Measurement, 51*(3), 755-765.
- Emmer, E. T., & Stough, L. M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist, 36*(2), 103-112.

- Evans, E. D., & Tribble, M. (1986). Perceived teaching problems, self-efficacy, and commitment to teaching among preservice teachers. *The Journal of Educational Research*, 80(2), 81-85.
- Evans, I. M., Harvey, S. T., Buckley, L., & Yan, E. (2009). Differentiating classroom climate concepts: Academic, management, and emotional environments. *Kōtuitui: New Zealand Journal of Social Sciences Online*, 4(2), 131-146.
- Evertson, C. M., & Neal, K. W. (2006). Looking into Learning-Centered Classrooms: Implications for Classroom Management. Working Paper. *National Education Association Research Department*.
- Evertson, C. M., & Weinstein, C. S. (2013). History of research on classroom management. In *Handbook of classroom management* (pp. 27-54): Routledge.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59-109.
- Freiberg, H. J., Huzinec, C. A., & Templeton, S. M. (2009). Classroom management—a pathway to student achievement: A study of fourteen inner-city elementary schools. *The Elementary School Journal*, 110(1), 63-80.
- Gallagher, H. A. (2004). Vaughn Elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education*, 79(4), 79-107.
- Glazerman, S., Mayer, D., & Decker, P. (2006). Alternative routes to teaching: The impacts of Teach for America on student achievement and other outcomes. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*, 25(1), 75-96.

- Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal*, 37(2), 479-507.
- Goe, L. (2007). The Link between Teacher Quality and Student Outcomes: A Research Synthesis. *National Comprehensive Center for Teacher Quality*.
- Goe, L., Bell, C., & Little, O. (2008). Approaches to Evaluating Teacher Effectiveness: A Research Synthesis. *National Comprehensive Center for Teacher Quality*.
- Goe, L., & Stickler, L. M. (2008). Teacher Quality and Student Achievement: Making the Most of Recent Research. TQ Research & Policy Brief. *National Comprehensive Center for Teacher Quality*.
- Goldhaber, D., & Anthony, E. (2003). Teacher Quality and Student Achievement. Urban Diversity Series.
- Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129-145.
- Goldring, R., Taie, S., & Riddles, M. (2014). Teacher Attrition and Mobility: Results from the 2012-13 Teacher Follow-Up Survey. First Look. NCES 2014-077. *National Center for Education Statistics*.
- Good, T. L., & Lavigne, A. L. (2017). *Looking in classrooms*: Routledge.
- Grossman, P., Loeb, S., Cohen, J., & Wyckoff, J. (2013). Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores. *American Journal of Education*, 119(3), 445-470.

- Guardino, C. A., & Fullerton, E. (2010). Changing behaviors by changing the classroom environment. *Teaching Exceptional Children, 42*(6), 8-13.
- Guo, Y., Piasta, S. B., Justice, L. M., & Kaderavek, J. N. (2010). Relations among preschool teachers' self-efficacy, classroom quality, and children's language and literacy gains. *Teaching and Teacher Education, 26*(4), 1094-1103.
- Guskey, T. R. (1982). Differences in teachers' perceptions of personal control of positive versus negative student learning outcomes. *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.
- Guskey, T. R., & Sparks, D. (2004). Linking professional development to improvements in student learning. *Research linking teacher preparation and student performance*, 11-21.
- Hagen, K. M., Gutkin, T. B., Wilson, C. P., & Oats, R. G. (1998). Using vicarious experience and verbal persuasion to enhance self-efficacy in pre-service teachers: "Priming the pump" for consultation. *School Psychology Quarterly, 13*(2), 169.
- Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child development, 76*(5), 949-967.
- Haney, W., Madaus, G., Kreitzer, A. (1987). Chapter 5: Charms Talismanic: Testing Teachers for the Improvement of American Education. *Review of research in education, 14*(1), 169-238.
- Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education review, 30*(3), 466-479.

- Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality*. Retrieved from
- Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, *100*(2), 267-271.
- Hardré, P. L., Huang, S. H., Chen, C. H., Chiang, C. T., Jen, F. L., & Warden, L. (2006). High school teachers' motivational perceptions and strategies in an East Asian Nation. *Asia-Pacific Journal of Teacher Education*, *34*(2), 199-221.
- Harris, D. N., Ingle, W. K., & Rutledge, S. A. (2014). How teacher evaluation methods matter for accountability: A comparative analysis of teacher effectiveness ratings by principals and teacher value-added measures. *American Educational Research Journal*, *51*(1), 73-112.
- Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*: routledge.
- Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication monographs*, *76*(4), 408-420.
- Hicks, S. (2012). Self-efficacy and classroom management: A correlation study regarding the factors that influence classroom management.
- Higgins, J. W., Williams, R. L., & McLaughlin, T. (2001). The effects of a token economy employing instructional consequences for a third-grade student with learning disabilities: A data-based case study. *Education and treatment of children*, 99-106.
- Hoffman, J. (2002). Flexible grouping strategies in the multiage classroom. *Theory into practice*, *41*(1), 47-52.



- Holden, G., Moncher, M. S., Schinke, S. P., & Barker, K. M. (1990). Self-efficacy of children and adolescents: A meta-analysis. *Psychological Reports, 66*(3), 1044-1046.
- Holsti, L., & Barr, R. G. (2006). The importance of electronic diary studies for occupational therapy research. *OTJR: Occupation, Participation and Health, 26*(2), 80-83.
- 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.
- Hsu, T. c. (2005). Research methods and data analysis procedures used by educational researchers. *International Journal of Research & Method in Education, 28*(2), 109-133.
- Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of labor Economics, 26*(1), 101-136.
- Jennings, J., & Rentner, D. S. (2006). Ten big effects of the No Child Left Behind Act on public schools. *Phi Delta Kappan, 88*(2), 110-113.
- Jimmieson, N. L., Hannam, R. L., & Yeo, G. B. (2010). Teacher organizational citizenship behaviours and job efficacy: Implications for student quality of school life. *British journal of Psychology, 101*(3), 453-479.
- Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education review, 27*(6), 615-631.
- Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources, 46*(3), 587-613.

- Kaplan, D. S., Peck, B. M., & Kaplan, H. B. (1997). Decomposing the academic failure–dropout relationship: A longitudinal analysis. *The Journal of Educational Research*, 90(6), 331-343.
- Keidel, S. A. (2014). *Teacher demographics, professional preparation, and training needs associated with classroom management based on teachers' self-reported survey*. Azusa Pacific University.
- Kennedy, M. M. (2010). Attribution error and the quest for teacher quality. *Educational researcher*, 39(8), 591-598.
- Kerr, M. M., & Nelson, C. M. (2002). *Strategies for addressing behavior problems in the classroom*: Merrill Upper Saddle River, NJ.
- Kihlstrom, J. F., Eich, E., Sandbrand, D., & Tobias, B. A. (2000). Emotion and memory: Implications for self-report. *The science of self-report: Implications for research and practice*, 81-99.
- Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education*, 79(4), 54-78.
- Klassen, R. M., Tze, V. M., Betts, S. M., & Gordon, K. A. (2011). Teacher efficacy research 1998–2009: Signs of progress or unfulfilled promise? *Educational Psychology Review*, 23(1), 21-43.
- Klem, A. M., & Connell, J. P. (2004). Relationships matter: Linking teacher support to student engagement and achievement. *Journal of school health*, 74(7), 262-273.

- Ku, K. Y., Ho, I. T., Hau, K.-T., & Lai, E. C. (2014). Integrating direct and inquiry-based instruction in the teaching of critical thinking: an intervention study. *Instructional Science, 42*(2), 251-269.
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology, 105*(3), 805.
- Kyriakides, L., Campbell, R., & Christofidou, E. (2002). Generating criteria for measuring teacher effectiveness through a self-evaluation approach: A complementary way of measuring teacher effectiveness. *School effectiveness and school improvement, 13*(3), 291-325.
- Lachman, H. (2006). An overview of the genetics of substance use disorders. *Current Psychiatry Reports, 8*(2), 133-143.
- Lavay, B., Henderson, H., French, R., & Guthrie, S. (2012). Behavior management instructional practices and content of college/university physical education teacher education (PETE) programs. *Physical Education & Sport Pedagogy, 17*(2), 195-210.
- LePage, P., Darling-Hammond, L., Akar, H., Gutierrez, C., Jenkins-Gunn, E., & Rosebrock, K. (2005). Classroom Management.
- Little, O., Goe, L., & Bell, C. (2009). A Practical Guide to Evaluating Teacher Effectiveness. *National Comprehensive Center for Teacher Quality*.
- MacSuga, A. S., & Simonsen, B. (2011). Increasing teachers' use of evidence-based classroom management strategies through consultation: Overview and case studies. *Beyond Behavior, 20*(2), 4-12.

- Marshall, J. C., & Smart, J. B. (2013). Teachers' transformation to inquiry-based instructional practice. *Creative Education, 4*(02), 132.
- Marzano, R. J., Marzano, J. S., & Pickering, D. (2003). *Classroom management that works: Research-based strategies for every teacher*: ASCD.
- Mathis, W. J., & Trujillo, T. M. (2016). Lessons from NCLB for the Every Student Succeeds Act. *National Education Policy Center*.
- Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis, 21*(1), 29-45.
- McCaffrey, D. F., Lockwood, J., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating Value-Added Models for Teacher Accountability*. Monograph: ERIC.
- McIntyre, T., & McIntyre, S. (2011). The teacher stress diary–TSD. *Unpublished instrument, Houston, TX: Texas Institute for Measurement, Evaluation, and Statistics, University of Houston*.
- McIntyre, T., McIntyre, S., & Durand, A. (2010). Teacher socio-demographic and professional survey. *Unpublished instrument, Houston, TX: Texas Institute for Measurement, Evaluation, and Statistics, University of Houston*.
- McIntyre, T. M., McIntyre, S. E., Barr, C. D., Woodward, P. S., Francis, D. J., Durand, A. C., . . . Kamarck, T. W. (2016). Longitudinal study of the feasibility of using ecological momentary assessment to study teacher stress: Objective and self-reported measures. *Journal of Occupational Health Psychology, 21*(4), 403.
- Medley, D. M., & Coker, H. (1987). The accuracy of principals' judgments of teacher performance. *The Journal of Educational Research, 80*(4), 242-247.

- Melser, N. A. (1999). Gifted students and cooperative learning: A study of grouping strategies. *Roeper Review*, 21(4), 315.
- Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Change in teacher efficacy and student self- and task-related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology*, 81(2), 247.
- Milanowski, A. T., Kimball, S. M., & White, B. (2004). The Relationship Between Standards-Based Teacher Evaluation Scores and Student Achievement: Replication and Extensions at Three Sites Consortium for Policy Research in Education (CPRE)-University of Wisconsin Working Paper Series. *TC*, 4(01).
- Mohamadi, F. S., & Asadzadeh, H. (2012). Testing the mediating role of teachers' self-efficacy beliefs in the relationship between sources of efficacy information and students achievement. *Asia Pacific Education Review*, 13(3), 427-433.
- Mojavezi, A., & Tamiz, M. P. (2012). The Impact of Teacher Self-efficacy on the Students' Motivation and Achievement. *Theory & Practice In Language Studies*, 2(3).
- Moritz, S. E., Feltz, D. L., Fahrbach, K. R., & Mack, D. E. (2000). The relation of self-efficacy measures to sport performance: A meta-analytic review. *Research quarterly for exercise and sport*, 71(3), 280-294.
- 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.
- Moskowitz, D. S., & Young, S. N. (2006). Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology. *Journal of Psychiatry and Neuroscience*, 31(1), 13.

- Motoca, L. M., Farmer, T. W., Hamm, J. V., Byun, S.-y., Lee, D. L., Brooks, D. S., . . . Moohr, M. M. (2014). Directed consultation, the SEALS Model, and teachers' classroom management. *Journal of Emotional and Behavioral Disorders, 22*(2), 119-129.
- Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of counseling psychology, 38*(1), 30.
- Murnane, R. J. (1975). The impact of school resources on the learning of inner city children.
- No Child Left Behind: Elementary and Secondary Education Act, (2001).
- Newton, X. A., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *education policy analysis archives, 18*, 23.
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237-257.
- Oliver, R. M., & Reschly, D. J. (2007). Effective Classroom Management: Teacher Preparation and Professional Development. TQ Connection Issue Paper. *National Comprehensive Center for Teacher Quality*.
- Oliver, R. M., & Reschly, D. J. (2010). Special education teacher preparation in classroom management: Implications for students with emotional and behavioral disorders. *Behavioral Disorders, 35*(3), 188-199.
- Orthner, D. K., Akos, P., Rose, R., Jones-Sanpei, H., Mercado, M., & Woolley, M. E. (2010). CareerStart: A middle school student engagement and academic achievement program. *Children & Schools, 32*(4), 223-234.
- Paine, S. C., Radicci, J., Rosellini, L.C., Deutchman, L. & Darch, C.B. (1983). *Structuring your classroom for academic success*. Champaign, IL: Research Press.

- Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology*, 86(2), 193.
- Palardy, G. J., & Rumberger, R. W. (2008). Teacher effectiveness in first grade: The importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis*, 30(2), 111-140.
- Park, D.-Y., & Logsdon, C. (2015). Effects of modeling instruction on descriptive writing and observational skills in middle school. *International Journal of Science and Mathematics Education*, 13(1), 71-94.
- Partee, G. L. (2012). *Using Multiple Evaluation Measures to Improve Teacher Effectiveness*. Retrieved from
- Partin, T. C. M., Robertson, R. E., Maggin, D. M., Oliver, R. M., & Wehby, J. H. (2009). Using teacher praise and opportunities to respond to promote appropriate student behavior. *Preventing School Failure: Alternative Education for Children and Youth*, 54(3), 172-178.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational researcher*, 38(2), 109-119.
- Pianta, R. C., Hamre, B. K., & Allen, J. P. (2012). Teacher-student relationships and engagement: Conceptualizing, measuring, and improving the capacity of classroom interactions. In *Handbook of research on student engagement* (pp. 365-386): Springer.

- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2006). CLASS: Classroom assessment scoring system manual preschool (Pre-K) version. In: Charlottesville, VA: Center for Advanced Study of Teaching and Learning. [www.virginia.edu/vprgs/CASTL](http://www.virginia.edu/vprgs/CASTL).
- Podolsky, A., Kini, T., Bishop, J., & Darling-Hammond, L. (2017). Sticky schools: How to find and keep teachers in the classroom. *Phi Delta Kappan*, 98(8), 19-25.
- Popham, W. J. (2006). *Assessment for educational leaders*: Pearson/Allyn and Bacon.
- Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational researcher*, 31(7), 3-14.
- Porter, A. C., & Brophy, J. (1988). Synthesis of research on good teaching: Insights from the work of the Institute for Research on Teaching. *Educational Leadership*, 45(8), 74-85.
- Porter, A. C., Kirst, M. W., Osthoff, E., Smithson, J., & Schneider, S. (1993). Reform up close: A classroom analysis. *Madison, WI: Wisconsin Center for Research in Education*.
- Praetorius, A.-K., Lenske, G., & Helmke, A. (2012). Observer ratings of instructional quality: Do they fulfill what they promise? *Learning and instruction*, 22(6), 387-400.
- Preacher, K. J., & Leonardelli, G. J. (2001). Calculation for the Sobel test. Retrieved January, 20, 2009.
- Qu, Y., & Becker, B. J. (2003). Does Traditional Teacher Certification Imply Quality? A Meta-Analysis.
- Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology*, 104(3), 700.
- Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes*: ERIC.



- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.
- Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: evidence for an accessibility model of emotional self-report. *Psychological bulletin*, 128(6), 934.
- Rockoff, J. E., & Speroni, C. (2010). Subjective and objective evaluations of teacher effectiveness. *American Economic Review*, 100(2), 261-266.
- Rose, J. S., & Medway, F. J. (1981). Measurement of teachers' beliefs in their control over student outcome. *The Journal of Educational Research*, 74(3), 185-190.
- Ross, J. A. (1992). Teacher efficacy and the effects of coaching on student achievement. *Canadian Journal of Education/Revue canadienne de l'education*, 51-65.
- Ross, J. A. (1998). The antecedents and consequences of teacher efficacy. *Advances in Research on Teaching*, 7, 49-74.
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80(1).
- Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale, survey research tells us about teacher effects on student achievement: Insights from the prospectus study of elementary schools.
- Rubeck, M., & Enochs, L. (1991). *A path analytic model of variables that influence science and chemistry teaching self-efficacy and outcome expectancy in middle school science teachers*. Paper presented at the National Association for Research in Science Teaching, Lake Geneva, WI.

- Saft, E. W., & Pianta, R. C. (2001). Teachers' perceptions of their relationships with students: Effects of child age, gender, and ethnicity of teachers and children. *School Psychology Quarterly, 16*(2), 125.
- Saklofske, D., Michayluk, J., & Randhawa, B. (1988). Teachers' efficacy and teaching behaviors. *Psychological Reports, 63*(2), 407-414.
- Schumacher, G., Grigsby, B., & Vesey, W. (2015). Determining effective teaching behaviors through the hiring process. *International Journal of Educational Management, 29*(1), 139-155.
- Shiffman, S. (2000). Real-time self-report of momentary states in the natural environment: Computerized ecological momentary assessment. *The science of self-report: Implications for research and practice, 277-296*.
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*.
- Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education and treatment of children, 351-380*.
- Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology, 99*(3), 611.
- Soodak, L. C., & Podell, D. M. (1993). Teacher efficacy and student problem as factors in special education referral. *The Journal of Special Education, 27*(1), 66-81.
- Stage, S. A., & Quiroz, D. R. (1997). A meta-analysis of interventions to decrease disruptive classroom behavior in public education settings. *School Psychology Review*.

- Staiger, D. O., & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *Journal of Economic perspectives*, 24(3), 97-118.
- Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological bulletin*, 124(2), 240.
- Stone, A. A., Schwartz, J. E., Neale, J. M., Shiffman, S., Marco, C. A., Hickcox, M., . . . Cruise, L. J. (1998). A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of personality and social psychology*, 74(6), 1670.
- Stone, A. A., Shiffman, S. S., & DeVries, M. W. (1999). Ecological momentary assessment.
- Stough, L. (2006). The place of classroom management and standards in teacher education. In: NJ: Erlbaum.
- Stoughton, E. H. (2007). "How will I get them to behave?": Pre service teachers reflect on classroom management. *Teaching and Teacher Education*, 23(7), 1024-1037.
- Stronge, J. H. (2006). Teacher evaluation and school improvement: Improving the educational landscape. *Evaluating teaching: A guide to current thinking and best practice*, 2, 1-23.
- Stronge, J. H. (2007). *Qualities of effective teachers* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
- Stronge, J. H., Ward, T. J., & Grant, L. W. (2011). What makes good teachers good? A cross-case analysis of the connection between teacher effectiveness and student achievement. *Journal of Teacher Education*, 62(4), 339-355.
- Thomas, D. L., & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of personality and social psychology*, 59(2), 291.

- Thoonen, E. E., Slegers, P. J., Oort, F. J., Peetsma, T. T., & Geijsel, F. P. (2011). How to improve teaching practices: The role of teacher motivation, organizational factors, and leadership practices. *Educational administration quarterly*, 47(3), 496-536.
- Tomlinson, C. A. (2000). Differentiation of Instruction in the Elementary Grades. ERIC Digest.
- Tomlinson, C. A., & Moon, T. R. (2013). *Assessment and student success in a differentiated classroom*: ASCD.
- Tschannen-Moran, M., & Hoy, A. W. (2007). The differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and Teacher Education*, 23(6), 944-956.
- Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, 68(2), 202-248.
- Tucker, P. D., Stronge, J. H., Gareis, C. R., & Beers, C. S. (2003). The efficacy of portfolios for teacher evaluation and professional development: Do they make a difference? *Educational administration quarterly*, 39(5), 572-602.
- Van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal*, 47(2), 497-527.
- Vaughn, S., Bos, C. S., & Schumm, J. S. (1997). *Teaching Mainstreamed, Diverse, and At-Risk Students in the General Education Classroom*: ERIC.
- Wang, H., Hall, N. C., & Rahimi, S. (2015). Self-efficacy and causal attributions in teachers: Effects on burnout, job satisfaction, illness, and quitting intentions. *Teaching and Teacher Education*, 47, 120-130.
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63(3), 249-294.

- Waxman, H. C., Hilberg, R. S., & Tharp, R. G. (2004). Future directions for classroom observation research. *Observational research in US classrooms: New approaches for understanding cultural and linguistic diversity*, 266-277.
- Weiland, I. S., Hudson, R. A., & Amador, J. M. (2014). Preservice formative assessment interviews: The development of competent questioning. *International Journal of Science and Mathematics Education*, 12(2), 329-352.
- Wenglinsky, H. (2000). How teaching matters: Bringing the classroom back into discussions of teacher quality.
- Wesley, D. A., & Vocke, D. E. (1992). Classroom Discipline and Teacher Education.
- Wharton-McDonald, R., Pressley, M., & Hampston, J. M. (1998). Literacy instruction in nine first-grade classrooms: Teacher characteristics and student achievement. *The Elementary School Journal*, 99(2), 101-128.
- Wijnia, L., Loyens, S. M., van Gog, T., Deros, E., & Schmidt, H. G. (2014). Is there a role for direct instruction in problem-based learning? Comparing student-constructed versus integrated model answers. *Learning and instruction*, 34, 22-31.
- Wolters, C. A., & Daugherty, S. G. (2007). Goal structures and teachers' sense of efficacy: Their relation and association to teaching experience and academic level. *Journal of Educational Psychology*, 99(1), 181.
- Wood, R., & Bandura, A. (1989). Social cognitive theory of organizational management. *Academy of management Review*, 14(3), 361-384.
- Woolfolk, A. E., Rosoff, B., & Hoy, W. K. (1990). Teachers' sense of efficacy and their beliefs about managing students. *Teaching and Teacher Education*, 6(2), 137-148.

- Woolfolk Hoy, A., Hoy, W. K., & Kurz, N. M. (2008). Teacher's academic optimism: The development and test of a new construct. *Teaching and Teacher Education, 24*(4), 821-835.
- Zee, M., & Koomen, H. M. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research, 86*(4), 981-1015.
- Zhao, X., Lynch Jr, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of consumer research, 37*(2), 197-206.

Appendix A.

*School and Teacher Demographics (N = 202)*

School/Teacher	<i>n</i>	%	<i>M</i>	<i>SD</i>	Min.	Max.
District 1 (17 campuses)	143	70.8				
District 2 (5 campuses)	59	29.2				
Number of Students			1,062.50	346.06	432	1,726
Grade/Course Taught*						
6 <sup>th</sup>	82	40.6				
7 <sup>th</sup>	101	50.0				
8 <sup>th</sup>	85	42.1				
Language Arts	65	32.2				
Math	61	30.2				
Social Studies	34	16.8				
Science	56	27.7				
Average number of students per class			24.7	6.0	5	45
Total number of years as educator			12.5	9.1	1	47

\*some teachers taught more than one grade or course