THE EFFECTS OF A THREE-STEP INSTRUCTIONAL METHOD ON TEACHER SELF-EFFICACY

A Dissertation in Educational Psychology

by

David E. Favre

© 2017 David E. Favre

Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

May 2017
The dissertation of David E. Favre was reviewed and approved* by the following:

Stephanie L. Knight
Professor of Educational Psychology
Associate Dean for Undergraduate and Graduate Studies
Dissertation Adviser
Chair of Committee

Robert J. Stevens
Professor of Educational Psychology

Hoi K. Suen
Distinguished Professor of Educational Psychology

Bernard J. Badiali
Associate Professor of Education
Professor in Charge: Curriculum and Supervision Program

Peggy N. Van Meter
Associate Professor of Educational Psychology
Professor in Charge: Educational Psychology Program

*Signatures are on file in the Graduate School.
Abstract

The purpose of this research was to determine the effects of instructional modeling on novice teacher self-efficacy development. Self-efficacy is a key teacher variable associated with many positive educational outcomes, yet little is known about how the observation of instructional modeling affects teachers' development of their self-efficacy for the specific behaviors they will need to effectively perform in their future classrooms. This study integrated methodologies which were responsive to four theoretical concerns raised in the literature to more appropriately measure changes to novice teacher self-efficacy: 1) pre-service teachers primarily form their self-efficacy from their observations of modeled teaching behaviors, 2) self-efficacy formation must be measured at the appropriate level of specificity (i.e. – the new pedagogical skill being learned), 3) self-efficacy formation must be measured as soon as possible after the targeted behavior is observed, and 4) researchers must take frequent measurements to observe small incremental changes to teacher self-efficacy.

Six modeled teaching skills that are common to second language teacher education were presented to participants in a series of videos to determine their effects on participant self-efficacy appraisals for these specific skills and their relationship to participants' assessment of their larger instructional strategies self-efficacy. Results indicated that participants consistently scored higher on posttest measures of the six micro-appraisal items than on the pretest measures. Moderate to high effect sizes were observed for the six treatments, suggesting that the observation of teaching behaviors modeled on videos has positive practical effects on preservice teachers’ appraisals of their self-efficacy for specific teaching tasks. Results also indicated that only one micro-appraisal change score was significantly related to participants' instructional strategies self-efficacy change. However, the cumulative score of the six micro-appraisal
changes was found to be moderately correlated with instructional strategies self-efficacy change, suggesting that cumulative scores of conceptually congruent items may be more predictive than individual items.
## Table of Contents

List of Figures ......................................................................................................................... ix
List of Tables .......................................................................................................................... x
Acknowledgement .................................................................................................................. xi

Chapter 1: Introduction ........................................................................................................... 1
  Statement of Problem ............................................................................................................. 2
  Novice Teachers .................................................................................................................... 2
  Gap in the Literature ............................................................................................................ 3
Conceptual and Measurement Problems with Teacher Self-efficacy ............................... 4
Problems with Understanding Self-efficacy Development in Teacher Education .......... 6
Rationale ................................................................................................................................. 9
Theoretical Framework for Measuring Teacher Self-efficacy Development .................... 10
Model Observation Primacy ................................................................................................. 11
Measuring Teacher Self-efficacy Development at the Task Level .................................... 12
Timing of Teacher Self-efficacy Assessment ....................................................................... 15
Microgenetic Methods .......................................................................................................... 16
Research Questions ............................................................................................................... 17
Significance ............................................................................................................................. 18

Chapter 2: Review of the Literature ..................................................................................... 20
  Self-efficacy ......................................................................................................................... 20
  Sources of Self-efficacy ....................................................................................................... 24
  Mastery Experience ............................................................................................................ 25
  Model Observation ............................................................................................................. 26
  Research on Modeling ....................................................................................................... 28
  Repeated Modeling ............................................................................................................ 29
  Video Modeling .................................................................................................................. 30
  Social Persuasion ............................................................................................................... 33
Physiological and Affective Feedback ................................................................................. 34
Research Exploring the Sources of Teacher Self-efficacy .................................................. 34
Malleability of Self-efficacy ................................................................................................. 35
Self-efficacy’s Influence on Human Functioning ................................................................. 37
Conceptualizing and Measuring Teacher Self-efficacy ........................................41
Teacher Self-efficacy Development ....................................................................44
Longitudinal Studies on Teacher Self-efficacy Development ............................46
Teacher Self-efficacy’s Impact on Educational Outcomes ................................47
Determining Teacher Self-efficacy ...................................................................49
Strength ...........................................................................................................50
Magnitude .........................................................................................................50
Generality ...........................................................................................................51
Summary ............................................................................................................54
Chapter 3: Pilot Studies .........................................................................................56
Pilot Study 1 .........................................................................................................56
Materials .............................................................................................................57
Week 1 Videos ....................................................................................................57
Week 2 Videos ....................................................................................................58
Week 3 Videos ....................................................................................................58
Week 4 Videos ....................................................................................................59
Week 5 Videos ....................................................................................................60
Week 6 Videos ....................................................................................................60
Procedures ..........................................................................................................61
Instruments ..........................................................................................................62
TSES – Short Form ...............................................................................................62
Causal Attribution Item .......................................................................................63
CDSII ....................................................................................................................63
Micro-appraisal Item ............................................................................................63
Data Analysis Procedures ....................................................................................64
Results .................................................................................................................64
Research Question 1 ............................................................................................65
Research Question 2 ............................................................................................66
Research Question 3 ............................................................................................67
Changes for the Dissertation Study ....................................................................69
Pilot Study 2 .........................................................................................................71
Chapter 6: Discussion ........................................................................................................122
Discussion of Results ......................................................................................................122
Research Question 1 ......................................................................................................123
Research Question 2 ......................................................................................................124
Implications ....................................................................................................................125
Implications for Theory .................................................................................................126
Implications for Research ..............................................................................................127
Implications for Practice ...............................................................................................128
Limitations .....................................................................................................................129
Participant Characteristics ............................................................................................129
Small Sample Size ........................................................................................................129
Possible Response Bias .................................................................................................130
Study Design Characteristics ........................................................................................130
Fidelity of Treatment ......................................................................................................130
Conceptual Nature of the Treatment .............................................................................131
Suggestions for Future Research ....................................................................................132
Study Replication ...........................................................................................................132
Integration of Treatment into Course Design ................................................................133
Experimental Control ....................................................................................................133
Longitudinal Study .........................................................................................................134
Conclusion .....................................................................................................................134
References ......................................................................................................................137
Appendix A: Weekly Video Presentations ......................................................................158
Appendix B: Teachers Sense of Efficacy Scale (TSES) – Short Form ..........................161
Appendix C: Pilot Study 1: Causal Attribution Item and Revised Causal Dimension Scale
  (CDSII) ..........................................................................................................................162
Appendix D: Pilot Study 2: Items and Appraisal Scale ..................................................163
Appendix E: Recruitment Email to Course Instructors ..................................................167
Appendix F: Voluntary Participant Sign-up ....................................................................168
Appendix G: Informed Consent ......................................................................................169
Appendix H: Experimental Group: Three-Step Treatment and Assessment Phases ....171
List of Figures

Figure 1: The Cyclical Nature of Teacher Efficacy ................................................................. 45
Figure 2: Cohen’s d Formula ..................................................................................................... 110
Figure 3: Post Hoc Power Analysis to Determine Observed Power ..................................... 110
Figure 4: Sensitivity Power Analysis to Determine a Minimal Detectable Effect (MDE) .... 111
Figure 5: Outliers for Micro-appraisal Gains (z-scores) ....................................................... 115
Figure 6: Outliers for Instructional Strategies Self-efficacy Gains ........................................ 116
Figure 7: Histogram of z-score Micro-appraisal Changes ....................................................... 117
Figure 8: Histogram of Instructional Strategies Self-efficacy Change ................................... 117
Figure 9: Q-Q Plot of z-score Micro-appraisal Changes ......................................................... 118
Figure 10: Q-Q Plot of z-score Micro-appraisal Change ....................................................... 118
Figure 11: Scatter Plot of z-scores of Micro-appraisal Changes Plotted Against Instructional Strategies Self-efficacy Change ................................................................................ 119
Figure 12: Scatter Plot Testing Homoscedasticity ................................................................ 120
List of Tables

Table 1: Participation Rates ...........................................................................................................64
Table 2: Group Mean Self-efficacy Changes ..................................................................................66
Table 3: Group Causation Attribution Orientations ......................................................................68
Table 4: Multitrait-Multimethod Matrix Correlations .................................................................75
Table 5: Weekly Group Activities .................................................................................................95
Table 6: Test of Normality of Micro-Appraisal Data ....................................................................101
Table 7: Test Statistics ..................................................................................................................102
Table 8: Rank Comparisons .........................................................................................................102
Table 9: Descriptive Statistics for Micro-Appraisals of Targeted Teaching Behaviors ..............103
Table 10: TSES Baseline Descriptive Statistics ..........................................................................105
Table 11: TSES Subscale and Composite Reliability ....................................................................106
Table 12: Mean Instructional Strategy Self-efficacy Scores ..........................................................108
Table 13: Shapiro-Wilk Test of Normality ....................................................................................112
Table 14: Spearman’s rho Correlations ......................................................................................113
Table 15: TSES Instructional Strategies Subscale Items ..............................................................131
Acknowledgement

Funding for this dissertation research was provided through a Student Research Initiation Grant from the Pennsylvania State University’s College of Education.
Epigraph

Try not to become a man of success, but rather try to become a man of value. He is considered successful in our day who gets more out of life than he puts in. But a man of value will give more than he receives.

- Albert Einstein (Life, May 2, 1955, p. 64)
Chapter 1 - Introduction

Best practices in second language (L2) teacher education are informed by evidence-based research that addresses broad areas of teacher professional development (Johnson, 2009). To effectively prepare pre-service teachers, L2 teacher education programs must consider three wide-ranging areas: “(1) the content of L2 teacher education programs…; (2) the pedagogies that are taught in L2 teacher education programs…; and (3) the instructional forms of delivery through which content and pedagogies are learned” (p. 11). An underlying and unifying concept to this comprehensive framework is the development of teacher self-efficacy by pre-service teachers as they experience their L2 teacher education programs. Although no curricula can fully prepare pre-service teachers for what they will encounter throughout the course of their professional lives, L2 teacher education programs must engage in practices that develop teacher self-efficacy across a broad range of specific teacher tasks that will foster teachers’ beliefs in their capacity to adapt to future challenges (Darling-Hammond, 2006b; Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998).

Understanding how teacher efficacy may be promoted by L2 teacher education practices is no easy undertaking, but this knowledge is critical for developing pre-service teachers’ beliefs in their ability to effectively apply the knowledge base of their L2 teacher education to their future practice (Chacón, 2005; Goker, 2006). Currently, we have only a limited conception of how teachers’ form their efficacy during their teacher education experiences. Wheatley (2005) stated that “a central challenge for researchers interested in the use of teachers’ efficacy in teacher education is to identify teacher education practices that lead to changes in teacher efficacy which in turn support meaningful changes in actual teaching” (p. 762). This study responds to Wheatley’s challenge by providing a theoretical conceptualization and methodology for how teacher self-efficacy may be measured during their formation as pre-service teachers.
engage in specific knowledge and skill building activities offered by their L2 education programs.

**Statement of Problem**

Teacher efficacy research has widely accepted Bandura’s (1986) four sources of information for self-efficacy appraisal (mastery experiences, vicarious experiences, verbal persuasion, and physiological/affective arousal) as being applicable to judgments of teaching efficacy. Yet inexplicably, there has been a historical lack of attention to these sources or how teacher self-efficacy may be promoted within teacher education (Klassen, Tze, Betts, & Gordon, 2011). Studies examining the impact of professional development on teacher efficacy fail to isolate and examine the effects of the individual sources (Ross & Bruce, 2007). Previous research has universally measured changes to teacher self-efficacy through global scales administered pre- and post- intervention (i.e., treatment, teacher education, or professional development). Consequently, we know very little about how precise teacher education activities may foster teacher self-efficacy development. If we do not understand these specific processes, professional development curricular designs essentially become a "black box" of non-specific treatments situated between baseline and post-intervention measurements. This study will isolate specific teacher education experiences (i.e., modeling of L2 teacher pedagogy) which contribute to teacher efficacy formation, growth, and calibration.

**Novice teachers.** Novice teachers are pre-service teachers attending their teacher education program and new teachers just beginning their careers. Novice teachers experience challenges to their self-efficacy based on the novelty of learning unfamiliar pedagogical skills and transitioning to new unexpected teaching conditions (Eccles, Midgley, & Adler, 1984; Pajares, 1997; Ross, 1998; Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998). When learning new
skills, novice teachers experience a pattern where their teacher self-efficacy increases during skill acquisition, followed by an implementation dip when these skills are integrated into their professional practice (Tschannen-Moran & McNaster, 2009; Woolfolk Hoy & Burke Spero, 2005). If significant changes to teaching contexts are perceived, teachers are inclined to reevaluate their self-efficacy for their capacity to competently perform expected teaching behaviors (Favre & Knight, 2016; Ross, 1998; Tschannen-Moran et al., 1998).

**Gap in the literature.** There is a gap in our knowledge of how teachers develop their self-efficacy for specific behaviors they will need to effectively perform in their classrooms. The research on teacher self-efficacy does not disaggregate unique sources of information that teachers consider when appraising their teaching efficacy. Investigations into teacher self-efficacy development have universally taken an integrated approach to any treatment that teachers receive to increase their teaching efficacy. Rowan, Correnti, and Miller (2002) insist that we need to more fully understand the individual processes that may produce any changes to understand why some conditions are more effective than others. Klassen and his colleagues (2011) suggested that research would be more relevant to the practice of teacher education if researchers “renew their focus on understanding how teacher efficacy is fostered by teacher education programs, the study of which requires a focus on the sources of teacher efficacy and a clearer understanding of how efficacy change over time” (p. 40).

Teacher self-efficacy has become an important construct in teacher education, and teacher educators should continue to explore how these beliefs develop, what factors contribute to strong and positive teaching efficacy in varied domains, and how teacher education programs can help preservice teachers develop high teacher self-efficacy. (Pajares, 2002b)
This study will focus on how instructional modeling of teaching behaviors influences novice teacher self-efficacy development because vicarious experiences are thought to have strongest influence over individuals lacking significant mastery experience (i.e. – novice teachers) in the specific context of the new skill being learned (Bandura, 1986; Pajares, 1997; Tschannen-Moran & Woolfolk Hoy, 2007). This research will address an existing knowledge gap in the literature because little is known about how vicarious experiences influence the formation and development of teacher self-efficacy (Pajares, 2002b).

Research on the measurement of the sources of teaching efficacy remains scarce in the literature (Klassen et al., 2011) despite repeated calls for its investigation (Goddard, Hoy, & Woolfolk Hoy, 2004; Henson, 2002).

More research into important sources of efficacy information that would tap the relative weight of vicarious experiences, verbal persuasion, mastery experiences, physiological arousal and contextual factors would be of great value as we attempt to learn how to better train and equip teachers for their complex tasks. (Tschannen-Moran & Woolfolk Hoy, 2007, p. 954)

**Conceptual and measurement problems with teacher self-efficacy.** Despite the relative maturity of the research in the area of teacher efficacy, there has been little consideration for the study of its sources. Klassen and his associates (2011) contend that there are continuing problems with a lack of attention by researchers to the sources of teacher efficacy, as well as continuing measurement and conceptual problems. Researchers investigating vicarious modeling effects consistently fail to attend to theoretical guidance when conceptualizing and measuring this construct (Favre, n.d.; Usher & Pajares, 2009). Researchers invariably surveyed participants’ assessment of their self-efficacy well after the targeted modeled behavior was
performed. Measurements of teacher self-efficacy have been unable to isolate the effects of modeling from other sources (Ross & Bruce, 2007; Tschannen-Moran & Woolfolk Hoy, 2001). Reliable measures of teacher self-efficacy information sources have yet to be developed (Klassen et al., 2011) and consequently it remains difficult to parcel out their interactive effects (Ross & Bruce, 2008). Researchers frequently generated social comparison statements for survey items such as “Individual differences among teachers account for the wide variations in student achievement” (Gibson & Dembo, 1984), and “My students usually do better than other teachers’ students” (Morris, 2010; See also Liu and Wilson, 2010) for a similar conceptualizations in student vicarious experience measurement. Survey questions also ask for some evaluation of a hypothesized model’s attributes, suggesting a formative comparison status for a previously observed model and omitting any framing of a task specific context, “My favorite teachers had good social skills” (Anderson & Betz, 2001).

While these studies may add substantially to what we know about the outcomes of teachers’ efficacy appraisals when they consider their vicarious experience, it ignores the very source or genesis of the information that is provided by models in teachers’ environments and factors theoretically important to the observation of specifically modeled tasks. At this immediate task-level consideration of one’s capacity for future task performance, sources of efficacy provide the essential core information that an individual considers as evidence and the basis for meaningful input for their efficacy appraisals. Traditional measures of teacher efficacy do not assess teacher efficacy appraisals at the appropriate task level, nor are they administered within a reasonable time after the task was modeled in the teacher education environment. When surveyed by instrumentation that aggregates these tasks at higher-levels of categorical performance and well after exposure to modeling performances, pre-service teachers must
summarize their efficacy for arbitrary categories based on their reflection of distal comparative information.

Problem with understanding self-efficacy development in teacher education. A comprehensive study on teacher preparation research prepared for the U.S Department of Education (Center for the Study of Teaching and Policy, 2001) found limited research concerning teacher preparation with only 57 articles qualifying for their study. Their synthesis of research studies focused on five key aspects for improving teacher preparation: subject matter preparation, pedagogical preparation, field practicum experience, organizational policies, and alternative certification programs. Only two of the studies (Guyton, Fox, & Sisk, 1991; Wilson, 1996) examined teacher preparation using teacher self-efficacy as a theoretical framework.

A more contemporary review of the literature on teacher efficacy research noted an expansion of studies related to teacher preparation (Klassen et al., 2011). During the 12-year timeframe that was examined, 218 articles on teacher self-efficacy were identified of which 64 (29%) investigated the self-efficacy of pre-service teachers. Irrespective of this increased attention, researchers have failed to fully examine the sources of novice teacher self-efficacy development nor have they significantly advanced theory in this area. Consequently a foundation for translating theory into practice is deficient and any specific recommendations to teacher education programs would be premature until these sources are more fully explored.

The literature on teacher self-efficacy research has frequently reported the positive correlation between teachers’ participation in their teacher education or professional development programs and increases to their teaching efficacy (Darling-Hammond, Chung, & Frelow, 2002; Ingvarson, Meiers, & Beavis, 2005; Ross, 2014; Ross & Bruce, 2007; Woolfolk Hoy & Burke Spero, 2005). This correlation has also been noted in L2 teacher education, but
appears to be an area less studied by researchers in this discipline (Faez & Valeo, 2012; Sehlaoui & Albrecht, 2011). Unfortunately, these studies often suffer from similar methodological flaws such as asking teachers to report the effects of their teacher education program on their teaching efficacy (at the time of their program completion) almost three years after they had actually finished their programs (Faez & Valeo, 2012).

Difficulties associated with appropriately assessing self-efficacy have been persistent and pervasive throughout the broad-spectrum of self-efficacy research (Pajares, 1997) including the specific research domain of teacher efficacy (Klassen et al., 2011; Tschannen-Moran et al., 1998; Wheatley, 2005). These large scale studies have drawn attention to continuing conceptual and measurement problems that have significantly affected the validity of any meaningful interpretation of scores for traditional measures. Problems with assessment have occurred when researchers drifted considerably from theoretical conceptions of self-efficacy measurement while constructing their instrumentation and implementing their assessment procedures. Errors in assessment are typically found when researchers do not distinguish self-efficacy from other belief concepts (i.e. intentionality, self-esteem, locus of control, and outcome expectancy), fail to construct assessment items on an appropriate level of specificity, or do not consider temporal effects in their methods (Bandura, 1986, 1997, 2006a; Bong, 2006; Gist & Mitchell, 1992; Pajares, 1997; Wheatley, 2005).

Another factor which confounds our understanding of pre-service teacher efficacy formation in L2 teacher education contexts is that degree and certification awarding programs do not appear to explicitly address this concern in their curricula or program designs. If programs do not prioritize developing this very important teacher quality in their designs, how can they know if their student-teachers are developing an increased sense of teaching efficacy that has been
frequently associated with many positive academic outcomes (Ross, 1998; Tschannen-Moran et al., 1998)? Compounding this oversight of teacher education programs not clearly addressing this critical teacher developmental characteristic is the fact that teacher efficacy is nowhere mentioned in the *Standards for the Recognition of Initial TESOL Programs in P–12 ESL Teacher Education* (TESOL, 2010) jointly developed by Teachers of English to Speakers of Other Languages (TESOL) and the National Council for Accreditation of Teacher Education (NCATE). Ironically, the TESOL/NCATE standards do address the importance of teachers developing an understanding of their own students’ beliefs, but this is nestled within the broader context of the standard for “Culture as it affects English language learning” (p. 40) for students. If the beliefs of English language learners (ELLs) are important enough to warrant a standard with performance indicators, why are important teacher beliefs (self-efficacy) not given any attention by the TESOL/NCATE standards?

Novice L2 teachers beginning their careers are often placed in unique classroom settings with little or no support to further develop their teaching skills and to address changes in their teaching self-efficacies resulting from the unexpected realities of their new teaching contexts. Pre-service L2 teachers often complain that they do not get enough actual teaching practice in authentic classrooms (Brandt, 2006; Johnson, 1996) and feel unprepared to teach in the immensely varied teaching contexts of their in-service L2 teaching placements. Teaching practice experiences in their L2 teacher education programs are very frequently monitored by experienced teacher educators or practicing teachers that provide a sheltered teaching experience for pre-service teachers. While these conditions do not qualify as a bonafide “mastery experience” source of teacher efficacy information when novices appraise their teaching efficacy, opportunities for teaching practice can offer novices opportunities to receive feedback
on their teaching performance (verbal persuasion) from expert teachers which may be used to inform their efficacy judgments. Novice teachers’ reported perceptions of inadequate preparation are primarily directed toward the inability of their teacher education programs to promote their efficacy development, rather than knowledge building or skill development (Eslami & Fatahi, 2008; Faez & Valeo, 2012).

**Rationale**

Cognitive processing of learning events impact teacher self-efficacy (Bandura, 1986, 1997; Tschannen-Moran et al., 1998). Research methods in this dissertation study attempted to isolate novice teacher attention to, or "noticing" of, these modeling events which critically inform their efficacy appraisals for future skill performance. I theorize that novice teachers making effort attributions will attend more fully to the salient elements of video modeling of new pedagogy. This increased attention will add to the amount of social comparison information regarding performance of these skills for their cognitive processing, thus creating a greater potential to affect their self-efficacy to competently perform these skills. Accumulation of these micro-appraisals for specific teacher skills should be positively correlated with posttreatment measures of larger subscale teacher performance expectation areas (i.e. - instructional strategies, classroom management, & student engagement).

Video modeling will be used in this study because it offers an efficient way to present vicarious observation source information (Bandura, 2004; Charlop-Christy, Le, & Freeman, 2000). Individuals can extract information from observing models on video that instructs their skill development. Social comparative information is also conveyed which informs their self-efficacy (Bandura, 1986, 2004; Usher & Pajares, 2008).
Theoretical framework for measuring teacher self-efficacy development. Acquiring a practical understanding for how L2 teacher education programs may improve teacher self-efficacy requires researchers to examine the process of its actual development. Wheatley (2005) suggested that such an understanding could only be achieved through a research design that included the use of microgenetic methods and interpretation of data collected over the course of any treatment intervention. Wheatley further advised that researchers should study teacher education programs, courses, or individual classes as the unit of analysis in order to better identify the teacher education practices that contribute to meaningful changes in the efficacy structures of pre-service teachers. Gist and Mitchell (1992) proposed that in novel task contexts, self-efficacy is formed through an assessment of task requirements, and contextual resources and constraints. Pre-service teachers in TESOL teacher education programs encounter many novel teaching contexts because the target concept they are teaching (the L2 and its associated culture) is also the medium of their instruction. This novel condition exists regardless of any prior teaching experience (i.e. – in-service teachers wanting to add a certification or degree).

This study’s theoretical framework for the measurement of pre-service teacher efficacy formation in L2 teacher education programs will respond to the concerns raised in the literature and integrate the methodological recommendations (Gist & Mitchell, 1992; Wheatley, 2005) mentioned above. More specifically, this study’s proposed theoretical framework is constituted by the following principles: 1) pre-service teachers primarily form their efficacy from their observations of the modeled targeted behavior, 2) efficacy formation must be measured at the appropriate level of specificity (i.e. – the new pedagogical skill being learned), 3) efficacy formation must be measured as soon as possible after the targeted behavior is observed, and 4)
researchers must use microgenetic methods when making their observations of targeted modeled behaviors.

*Model observation primacy.* Tschannen-Moran and Woolfolk Hoy (2007) recommend more research into important sources of efficacy information to improve teacher education programs’ abilities to better prepare pre-service teachers. They stated that “inputs such as verbal persuasion, vicarious experiences, and emotional arousal may well be most salient for preservice teachers who lack significant mastery experiences” (p. 954). Eccles and colleagues (1984) found that observation of models may be a more significant source of efficacy information during transitional periods. Although their study evaluated students only, it also has theoretical application for pre-service and transitioning teachers. Bandura (1986) emphasized that vicarious experiences may be more dominant in the development of self-efficacy in the absence of mastery experience. Most L2 teacher education programs do not provide opportunities for pre-service teachers to engage in authentic mastery experiences early in their program. Many programs provide sheltered student-teaching internship experiences where expert teacher educators observe pre-service teachers as they student-teach. Teacher educators are consistently present to provide correction and guidance, either through explicit or implicit means. However, the elevated levels of self-efficacy that pre-service teachers often report during their sheltered experiences frequently plummet when pre-service teachers begin their professional practice in unsheltered contexts (Tchannen-Moran et al., 1998; Tschannen-Moran & Woolfolk Hoy, 2007; Woolfolk Hoy & Burke Spero, 2005). Additionally, even the most recognized TESOL certification programs: the Trinity College London Certificate in TESOL (Cert TESOL) and the Cambridge ESOL Certificate in English Language Teaching to Adults (CELTA) typically offer just 6 contact hours of supervised teaching practice to their pre-service teachers (Brandt, 2006).
Given the fleeting effects of verbal persuasion and teacher inattention to their emotional and physiological arousal states for their teacher efficacy appraisals, it is logical to look toward vicarious experiences of model observation as a primary source of information for pre-service and transitioning teachers. The potential for observational learning to impact the efficacy of inexperienced teachers having few mastery experiences, as well as experienced teachers facing transitional ambiguity in their teaching contexts, clearly calls for an examination of this source as a primary source of information for efficacy appraisal in L2 teacher education contexts.

At this point it is necessary to distinguish between vicarious learning from modeled behavior and how observation of models can inform self-efficacy appraisals. Vicarious learning from observation of modeled behavior has four components: attention, retention, reproduction, and motivation (Bandura, 1997). However, only the first two components are necessary for individuals to judge their efficacy for a specific task. Consequently, any efficacy formation measurement must consider if the modeled behavior was attended to and retained by pre-service teacher observers if we are to be able to build a case for the construct validity of this method.

**Measuring teacher self-efficacy development at the task level.** Bandura (1997, 2006a) found that instruments measuring self-efficacy were often constructed too broadly. He suggested finding an appropriate balance between generality and specificity for self-efficacy measures. If an instrument is too general, teachers may not be able to properly conceptualize the task category to provide an accurate assessment of their efficacy. If the instrument is too specific, then any meaningful generalization of teacher efficacy beyond the exact context of scale items would be difficult to justify. Considerations for the level of item specificity are rarely reported in validations of teacher efficacy instrumentation. Bandura has provided some guidance to researchers for striking the proper balance between generality and specificity. He advised
researchers that “the optimal level of generality at which self-efficacy is assessed varies depending on what one seeks to predict and the degree of foreknowledge of the situational demands” (1986, p. 49). Bandura (2006a) also warned against using global measures of perceived self-efficacy because items on such assessments reflect a great deal of ambiguity for what is being probed and much of the contextual references are obscured. Because most items have little relevance to the domain of functioning they purport to assess, such general measures of teacher efficacy have limited explanatory and predictive value.

The connection of self-efficacy appraisal to the appropriate level of task is critical to the construct validity of the interpretation of scores for an assessment instrument (Gist & Mitchell, 1992). What cognitive considerations do pre-service teachers make when they are asked to appraise their capacity to perform a task at a specific level? Pajares (1996) expanded on this concern for striking an appropriate balance between specificity and generality of item focus:

Omnibus tests that aim to assess general self-efficacy provide global scores that decontextualize the self-efficacy-behavior correspondence and transform self-efficacy into a generalized personality trait rather than the context-specific judgments Bandura suggests they are. Generalized self-efficacy instruments assess people's general confidence that they can succeed at tasks and in situations without specifying what these tasks or situations are. Even domain-specific omnibus measures are problematic if composite multiscale scores drawn from differing subsections of the domain are used. (p. 547)

Pajares (1996) similarly warned researchers about going to the other extreme when constructing efficacy measures when he summarized Lent and Hackett’s (1987) observation that “specificity and precision are often purchased at the expense of external validity and practical relevance.”
Bong (2006) also recommended determining the appropriate level of specificity when constructing self-efficacy assessment instrumentation. She suggests that finding the most suitable balance in item construction requires a careful analysis of the targeted task performance considered for prediction.

If the goal is to predict students’ levels of performance on some defined academic task, researchers must analyze the types of skills, knowledge, and potential constraints on performance involved in successful accomplishment of the task before creating a self-efficacy scale. (p. 296)

Unfortunately, finding the right level of performance for efficacy measurement has been elusive for researchers or beyond their considerations. “Perhaps the greatest challenge has to do with finding the appropriate level of specificity for measurement” (Tschannen-Moran et al., 1998, p. 219).

Typical scale items only probe teacher efficacy for general categories of pedagogical tasks. For example, the TES has two sub-scales (Personal Teaching Efficacy and General Teaching Efficacy) with items phrased to elicit teacher appraisals of their efficacy to perform in general contexts that aggregate specific teacher tasks: “When a student does better than usual, many times it is because I exerted a little extra effort” and “When the grades of my students improve it is usually because I found more effective teaching approaches” (Gibson & Dembo, 1984). Although the Teachers’ Sense of Efficacy Scale (TSES) provides a narrower focus on teacher efficacy through its use of three sub-scales (Efficacy in Student Engagement, Efficacy in Instructional Strategies, and Efficacy in Classroom Management), its specific items are also constructed to tap teachers’ appraisals of their efficacy to perform categorically rather than at the task level: “How much can you use a variety of assessment strategies?” and “How much can you
do to adjust your lessons to the proper level for individual students?” (Tschannen-Moran & Woolfolk Hoy, 2001). When pre-service teachers consider such items, they are forced to reflect on their ability to perform in each of the specific tasks that fall into the categorical level represented by the item and to aggregate their appraisals in their response. For example, when considering the item - “How much can you use a variety of assessment strategies?”, teachers must assess their capacity to use standardized written tests, use rubrics to assess students’ oral responses to teacher-made questions, use portfolio assessments, etc. This aggregated appraisal of teaching efficacy does not allow for teacher education programs to assess which specific pedagogical skill development will better promote improvement to pre-service teacher self-efficacy.

**Timing of teacher self-efficacy assessment.** The lack of immediacy when asking teachers to appraise their efficacy for a specific task also appears to be a weakness in traditional measurements. Bandura (1977) suggested that self-efficacy is more malleable during its formation than in later years, when the stronger effects of mastery experience contribute to its stability. This suggests that early formation of efficacy for novel tasks may vary widely as pre-service teachers reflect on the scant information provided to them on their ability to perform a new pedagogical task. But it is this information that is of most concern if we are to evaluate the effectiveness of teacher educators’ instructional practices in promoting teaching self-efficacy.

Wood and Bandura (1989) demonstrated that the manner in which individuals interpret the source of their ability can affect the stability of their self-efficacy when they encounter disconfirmatory evidence. Individuals who construed their ability as an acquired skill demonstrated a highly resilient sense of efficacy to perform challenging new tasks, while individuals who conceptualized their ability as being indicative of an inherited aptitude sharply
decreased the strength of their self-efficacy appraisals after they encountered difficulties in their task performance. An individual’s appraisal of their efficacy may change over time when new intervening information and experiences are considered.

The degree of change that might be expected in self-efficacy is partly a function of the variability and locus of its determinants; thus, individual differences in the composition of self-efficacy must be considered when interventions are designed to enhance it. (Gist & Mitchell, 1992, p. 200)

These findings suggest that efficacy for perceived capacity to perform a specific task may vary considerably if measured at different times when pre-service teachers base their efficacy appraisals on recent confirmatory or disconfirmatory evidence. Intervening events experienced by pre-service teachers may affect their self-efficacy assessments. It is therefore prudent to measure efficacy formation immediately after exposure to the source of information that is being considered (i.e. – immediately after observing the modeled targeted behavior).

**Microgenetic methods.** Wheatley’s (2005) suggestion for using the microgenetic method for understanding how efficacy is developed is appropriate because efficacy development is a cognitive developmental change which necessitates methods that produce data about the specific change in question. The microgenetic method involves three essential properties of observation:

(a) Observations span the entire period from the beginning of the change to the time at which it reaches a relatively stable state. (b) The density of observations is high relative to the rate of change of the phenomenon. (c) Observed behavior is subject to intensive trial-by-trial analysis, with the goal of inferring the processes that give rise to both the quantitative and qualitative aspects of change. (Siegler & Crowley, 1991, p. 606)
Use of the microgenetic method will yield data which may provide insight for how the observation of modeled teaching tasks can affect pre-service teachers’ cognitive development of their efficacy. Tschannen-Moran and her colleagues (1998) stated that “modeling is known to influence self-efficacy, but less is known about what types of information from observation are particularly useful in determining efficacy” (pp. 239-240). Close and systematic observations are necessary to yield data that will determine how teacher educators specifically model targeted pedagogical tasks that may affect pre-service teacher efficacy formation for their capacity to perform a particular behavior. The microgenetic method of observation is also theoretically consistent with the cyclical nature of teacher efficacy development proposed by Tschannen-Moran and her colleagues (1998). Observations should occur over the duration of the teacher education program experience as pre-service teachers form their efficacy for specific tasks.

**Research Questions**

This dissertation study responds to the concerns raised in the literature for measuring teacher self-efficacy change. These concerns are that not enough attention has been given to the sources of self-efficacy (Klassen et al., 2011), measurement methods frequently do not adequately consider the appropriate level of specificity of teaching behaviors (Bandura, 1997; Gist & Mitchell, 1992; Pajares, 1996), instrumentation is time sensitive and is frequently administered too long after the participant has experienced a self-efficacy informing event (Bandura, 1986; Hensen, 2012; Pajares & Miller, 1994) and data collection methods rely on infrequent measurement of participants as their efficacy is developing (Bandura, 1997; Wheatley, 2005).

This study examined the effects of video modeling on changes to pre-service teacher self-efficacy. Specifically, this study’s research questions are:
1. How does a three-step instructional method which includes: 1) reading the definition of novel skills, 2) viewing a direct instruction video about how to perform the skills, and 3) viewing three videos of models demonstrating the skills in authentic contexts affect the magnitude of teachers’ micro-appraisals of their self-efficacy for performing the modeled teaching behaviors?

2. What is the relationship between teachers’ micro-appraisals for specific modeled teaching behaviors and their overall self-efficacy assessment of the larger sub-scale to which these modeled behaviors represent? (e.g., How are teachers’ micro-appraisals for their ability to perform a recast of an English language learner’s speech error related to their assessment of their self-efficacy for performing instructional strategies?)

Significance

Findings from this research will provide new insight into teacher self-efficacy development and assessment within the specific context of L2 teacher education. A theoretical foundation is presented for a more precise method of conceptualizing and measuring pre-service teacher efficacy development which addresses ongoing specific theoretical and methodological concerns found in the literature. It is hoped that the the method for measuring pre-service teacher efficacy formation will bridge the gap between research and practice by generating meaningful recommendations for teacher educators to positively affect pre-service teacher efficacy development. If teacher self-efficacy is critical to the process of teaching, it should be made an explicit focus of teacher education programs (Pajares, 2002b).

Unfortunately, the body of research on teacher efficacy presents little utility for teacher education programs wanting to know how they can design their curricula, or encourage specific teacher educator pedagogical behaviors, to promote efficacy development for their pre-service
teachers. How then can research into teacher efficacy be more meaningfully applied to teacher education programs? Essentially, this question prompts us to look for ways to bridge the divide between research and practice in this area of teacher education.

With a better understanding of how to promote the development of teaching efficacy, teacher education programs will be better positioned to prepare pre-service teachers with a sense of agency to adapt to the challenges presented by the unique contexts of their impending L2 teaching placements. Continued acceptance of conventional methods for measurement of teaching self-efficacy will not provide us with any meaningful understanding for how teacher education programs can use specific practices to improve pre-service teacher self-efficacy. Without this explicit task-level efficacy formation knowledge, any changes to pre-service teacher efficacy by L2 teacher education programs cannot be adequately conceptualized or meaningfully measured. It is unproductive to continue to rely simply on a ‘black box theory’ for why teacher self-efficacy changed after any treatment intervention. We must examine these developmental changes as they occur if any meaningful recommendation are to be given to L2 teacher education programs to improve the self-efficacy of their teachers-in-training.
Chapter 2 – Review of the Literature

This review of the literature will examine social cognitive theory’s conceptions of self-efficacy, theoretical sources of self-efficacy, conditions that effect its malleability, and its influence on a variety of human functioning. Conceptualizations of teacher self-efficacy will then be reviewed with an emphasis on how teacher self-efficacy is thought to develop and have influence over educational outcomes. Findings from research on self-efficacy (Klassen et al., 2011; Zee & Koomen, 2016) will be considered to illuminate concerns for teacher self-efficacy measurement and relevance to practice.

Self-efficacy

Self-efficacy has been defined as an individual’s “beliefs about their capabilities to produce at designated levels of performance that exercise influence over events that affect their lives” (Bandura, 1994, p. 71). Self-efficacy is situated within social cognitive theory’s conceptualization of triadic reciprocal determinism where personal factors, behavior, and environmental factors mutually interact as determinants of human functioning (Bandura, 1986, 1989; Pajares, 2002a). An individual’s functioning can be understood by the mutual interactions of these three components. A key mediator of triadic reciprocal determinism is human agency or the ability of an individual to make choices. Self-efficacy is closely connected to human agency in that one’s beliefs about one’s capacity for learning and functioning at a designated level of behavior influences one’s choices regarding task engagement and persistence. Self-efficacy is a key construct of social-cognitive theory as it explains how people can achieve a sense of agency in their lives (Bandura, 1997). Self-efficacy is associated with an individual’s willingness to engage in a task, and the amount of effort and persistence an individual is willing to exert as they encounter difficulties while performing the task (Bandura, 1986).
Self-efficacy is a personal factor (i.e., an individual’s belief about their capacity to perform a specific behavior) that interacts reciprocally with behavioral, environmental, and other personal factors. An individual’s self-efficacy affects how they will alter their behavior, influence their environment, and interpret related cognitions. Conversely, due to the reciprocal nature of these interactions, an individual’s self-efficacy is also influenced by their appraisals of their behavior, perceptions of their environment, and the outcomes of self-reflective processes. Teachers’ self-efficacy judgments are partially informed by the cognitive appraisal of their past teaching behaviors. These judgments may consequently be used to adapt their teaching behavior to more effectively achieve desired outcomes. Teachers consider information derived from their teaching environment in their efficacy judgments. These evaluations may inform their choice of teaching environment or inform their behaviors to manipulate the conditions of their environmental contexts. Individual reflection allows teachers to reinterpret their self-efficacy as they consider the input from new experience.

Self-efficacy likewise “refers to beliefs in one’s capabilities to mobilize the motivation, cognitive resources, and courses of action needed to meet given situational demands” (Wood & Bandura, 1989, p. 408). This explanation emphasizes three important aspects of self-efficacy: its task specificity, dynamic nature, and mobilization component (Gist & Mitchell, 1992). First, a self-efficacy appraisal is a comprehensive assessment of perceived capability to competently perform a specific behavior. Teachers consider their beliefs, cognitions, and other personal characteristics, as well as environmental information when making efficacy judgments for their capacity to perform a specific teaching behavior. When assessing more generalized teaching capacities, teachers consider a wide array of capabilities they consider reflective of good teaching (Woolfolk Hoy & Burke Spero, 2005). Second, self-efficacy’s dynamic nature means
that it may change over time as new information is considered. “Execution of a skill must be constantly varied to suit changing circumstances” (Bandura, 1989, p. 1181). However, Bandura (1997) cautioned that “dislodging a low sense of personal efficacy requires explicit, compelling feedback that forcefully disputes the preexisting disbelief in one’s capabilities” (p. 82). Teachers may alter their efficacy judgments as they encounter and reflect on varying task conditions, and consider evidence, which either bolsters or contradicts their teaching efficacy. Although self-efficacy appear resistive to change once established, Bandura (1977b) suggested that self-efficacy may be changed more readily during early learning circumstances, as is the case in teacher education contexts. Third, individuals motivate themselves by setting progressively higher standards of achievement for successful task completion. They then mobilize their efforts based on their estimation of what it will take to achieve their desired results. Variation in task context will also require adaptive performance to continue functioning at a desired level of competence. Teachers must often adapt their pedagogy to address challenges to their teaching efficacy presented by expectations for implementing teaching innovations and responding to the ever changing contexts of their classrooms (i.e., the changing educational needs of their students). Teachers with high self-efficacy see these contextual alterations as positive conditions and frequently adopt a mastery orientation (Bandura, 1977a, 1986, 1994; Pajares, 2002a).

Self-efficacy is considered to be a future-oriented appraisal to perform a specific behavior which is informed by an individual’s perceptions of their competence rather than their actual competence (Bandura, 1977b, 1986; Woolfolk Hoy & Burke Spero, 2005). Bandura emphasized the distinction between an individual’s beliefs and their actions (performance) when he stated – “A capability is only as good as its execution. The self-assurance with which people approach and manage difficult tasks determines whether they make good or poor use of their capabilities.
Insidious self-doubts can easily overrule the best of skills” (Bandura, 1997, p. 35). Self-efficacy may forecast an individual’s potential capacity for successfully engaging in a future task, such as a teacher’s belief in the ability to effectively use a newly learned teaching strategy (Southerland, Sowell, Blanchard, & Granger, 2011). The future orientation of self-efficacy centers on an individual’s estimation of their capacity and not their intention. Consequently, items on measures of self-efficacy should be phrased using *can* rather than *will* (Bandura, 2006a).

Social cognitive theory distinguishes self-efficacy as an ability expectation. An individual’s efficacy expectancy reflects their belief that they can orchestrate the essential actions to competently perform a given task. This differs from social cognitive theory’s conception of outcome expectancy. An individual’s outcome expectancy is their belief in the likelihood that successfully completing a task will lead to certain consequences (Bandura, 1977b, 1986, 1997; Tschannen-Moran et al., 1998). Self-efficacy precedes and has causal influence over outcome expectancy in social cognitive theory (Bandura, 1978, 1997, 2006b). Bandura (1997) emphasized the causal effect of self-efficacy on outcome expectations when he asserted:

> The outcomes people anticipate depend largely on their judgments of how well they will be able to perform in given situations. To claim…that people visualize outcomes and then infer their own capabilities from the imagined outcomes is to invoke a peculiar system of backward causation in which the outcomes that flow from actions are made to precede the actions. (p. 21)

Bandura emphasized that outcome expectancies add little to the predictive power of self-efficacy measures because expected outcomes emanate more from an individual’s belief in their level of competency to perform a given task than from their calculation that a consequence will predictably follow accomplishing a task at a desired level (Bandura, 1986, 1997; Tschannen-
Moran et al., 1998). Physical, social, and self-evaluative outcomes that are positive serve as incentives and those that are negative serve as disincentives. An individual’s anticipation of obtaining a given outcome is mostly dependent on their appraisal of their expected performance level (Bandura, 1994, 2006a).

Self-efficacy is often confused with other psychological conceptions featuring a self-orientation but it is conceptually distinct from constructs such as self-concept, self-worth, and self-esteem (Tschannen-Moran et al., 1998; Schunk, 1991). Bandura (1997) also stressed that self-efficacy differs from our everyday understanding of the term “confidence” since a self-efficacy assessment “includes both an affirmation of a capability level and the strength of that belief” while “confidence is a nondescript term that refers to strength of belief but does not necessarily specify what that certainty is about” (p. 382). However, Bandura (2006a) later contradicted his position with his frequent use of confidence as a proxy metric for self-efficacy in his widely cited guide for constructing self-efficacy scales. Other popular instruments in teacher efficacy assessment use different phrasing to get at teachers’ self-efficacy such as asking respondents to “indicate the degree to which you agree or disagree with each statement” (Gibson & Dembo, 1984), rate “How much can you do to…” (Tschannen-Moran & Woolfolk Hoy, 2001), and “How certain are you that you can…” (Skaalvik & Skaalvik, 2007). Some care in phrasing should be taken to more accurately capture an individual’s assessment of their self-efficacy. Bandura stated that “items should be phrased in terms of can do rather than will do. Can is a judgment of capability; will is a statement of intention.” (Bandura, 2006a, p. 308). It appears that the temporal aspect of self-efficacy is more critical to maintain than slight semantic variations that may not be meaningfully considered when individuals appraise their self-efficacy.

**Sources of Self-efficacy**
Bandura (1977b, 1986) proposed four sources of information that individuals may consider when appraising their self-efficacy (mastery experiences, vicarious experiences, verbal persuasion, and states of affective and physiological arousal). These sources have been widely accepted by the research community without significant evaluation or challenge (Klassen, et al., 2011).

**Mastery experience.** Mastery experiences were hypothesized to have the strongest influence over self-efficacy development (Bandura, 1997). Successful task performance can strengthen an individual’s sense of self-efficacy, while task failure can damage or weaken it. However, individuals who experience frequent success on easily achieved tasks may come to expect quick results and become discouraged when they fail. Mastery experience in tasks where obstacles are present and overcome can foster a resilient sense of efficacy (Bandura, 1994). Research in teacher education has generally supported this hypothesis with regards to teachers beyond their preservice education. Tschannen-Moran and Woolfolk Hoy (2007) found that satisfaction with past professional performances was a stronger predictor for self-efficacy among teachers who have been teaching for some time than it was for their novice counterparts. Teachers who have continued their education and obtained higher-level teaching degrees often assess their teaching self-efficacy more strongly than those with lesser degrees (Hoy & Woolfolk, 1993; Williams, 2009). Teachers judge their capacity to positively affect student outcomes by reflecting on their past mastery experiences, including the mastery experiences during their student-teaching and induction year (Woolfolk Hoy & Burke Spero, 2005). But, are all teaching contexts comparable in regards to forming mastery experiences? Tschannen-Moran and her colleagues (1998) asserted:
Only in a situation of actual teaching can an individual assess the capabilities she or he brings to the task and experience the consequences of those capabilities. In situations of actual teaching, teachers gain information about how their strengths and weaknesses play out in managing, instructing, and evaluating a group of students. (p. 229)

Unfortunately, these authors did not elaborate on what constitutes a situation of actual teaching, but did suggest that student-teachers frequently misjudge the real demands associated with teaching while in their teaching practicums. Errors in judgement during student-teaching may contribute to preservice teachers’ difficulty with accurately calibrating their self-efficacy (Pajares, 1997).

**Model observation.** Bandura (1986, 1997) asserted that individuals who lack mastery experiences are more influenced by vicarious observation of models in their environment when making self-efficacy judgements. Though less powerful than mastery experience, vicarious observation of models remain a significant source of information for self-efficacy assessment. “Seeing people similar to oneself succeed by sustained effort raises observers' beliefs that they too possess the capabilities to master comparable activities to succeed” (Bandura, 1994). Model observation can assist individuals in their understanding of subject matter content, behavioral processes, and the consequences of behaviors. Model observation also informs self-efficacy through means of social comparison (Schunk, 2011). Models may be live persons, such as teachers or peers; or they may be symbolically represented in videos, photographs or graphics (Bandura, 1977b). "Because most performances are evaluated in terms of social criteria, social comparative information figures prominently in self-efficacy appraisals" (Bandura, 1986, p. 400). Perceived model similarity strongly mediates the influence that modeling has on an individual’s perceived self-efficacy (Bandura, 1986). Positive comparison to models who
competently perform a task can lead to a comparative assessment easily understood in terms of an “if they can do it, so can I” approach to performance. Similarly, negative comparisons to models who succeed, or positive comparisons to models who fail, may lead to appraisals which undermine an individual’s self-efficacy for the modeled task (Bandura, 1994). However, perceived model similarity is not the sole consideration of individuals during observation, “people are not about to discard information that makes them more efficacious just because it comes from and dissimilar source” (Bandura, 1997, p. 101). Individuals will also seek out expert models who demonstrate task competence in areas where they have ambition to succeed. When this is the case, individuals will pursue models whose differences demonstrate high levels of task competence such as those with high status, prestige, and power (Bandura, 1997). Additionally, self-efficacy appraisals are not solely based on task outcomes. Brown and Inouye (1978) found that participants’ self-efficacy was negatively affected when they observed a model who failed at an anagram task if they perceived themselves as having comparable ability. Participants who judged their capability as being superior to the model’s capability were not negatively affected. Vicarious learning has some advantages over enactive learning in that there are more opportunities for learning provided by observing competently modeled behaviors and individuals who learn vicariously are insulated from directly experiencing any negative consequences for task failure (Schunk, 2011).

Learning would be exceedingly laborious, not to mention hazardous, if people had to rely solely on the effects of their own actions to inform them what to do. Fortunately, most human behavior is learned observationally through modeling: from observing others one forms an idea of how new behaviors are performed, and on later occasions this coded information serves as a guide for action. (Bandura, 1977b, p. 22)
While an individual is more likely to reproduce a behavior that was successfully modeled, rewarded or performed by a positively compared model; an individual’s agency for choosing or not choosing to reproduce a modeled behavior or modifying the observed behavior should also be considered when determining if vicarious learning has occurred (i.e. – the behavior need not be imitated or reproduced to demonstrate learning).

**Research on modeling.** Modeling is often cited as an important component of effective practices in pre-service teacher education (Darling-Hammond, 2006b; Korthagen, 2004; Supovitz & Turner, 2000). Brownell, Ross, Colón, and McCallum (2005) reviewed two large-scale studies sponsored by the Association of American Colleges of Teacher Education (AACTE) and the International Reading Association (IRA) that examined 15 institutions recognized for their excellence in teacher education. Along with other critical program features, they found that faculty used the modeling of active pedagogy to promote a “conscious blending of theory, disciplinary knowledge, subject-specific pedagogical knowledge and practice” which helped their students to link these coursework features to their own teaching practices. Grossman’s (1987) review of research on pedagogical methods in teacher education found that preservice teachers’ observations of their teacher educator’s modeling of teaching skills could be as effective as a mastery experience for developing their confidence for using those skills.

Modeling in teacher education had also been shown to improve student-teachers development of practices that support their ability to engage in effective teaching behaviors. Loughran (1996) found that his modeling of reflective teaching practices, promoted a sense of confidence in his student-teachers to use their developing teaching skills. He modeled reflective teaching practices by using a think aloud approach when he encountered a teaching problem and by sharing his own teaching journal reflections. His follow-up interviews with his students
revealed that they were aware of the purpose of his modeling and they expressed an interest in using similar reflective practices in their future teaching. Smith, Frey and Tollefson (2003) found that modeling conducted by teacher educators made positively affected preservice teachers’ understanding of teacher collaboration and their beliefs about engaging in successful collaborative efforts. Preservice teachers were significantly affected by their observations of their teacher educators working together to effectively meet their educational needs.

Repeated modeling. Studies of repeated modeling in teacher education note improvements for retention and transfer of teaching skills. Singer, Lotter, Feller and Gates (2011) studied the impact of a professional development program on the ability of middle school science teachers to use inquiry-based pedagogical practices. They found that teachers credited their observations of the repeated modeling of their professional development instructors as being influential in their learning of these new pedagogical skills. The authors suggest that consistent modeling needs to occur across varying teaching contexts to promote retention and transfer. Kropiewnicki (2006) also found that repeated modeling of instructional strategies for developing children’s literacy benefited her pre-service teachers’ skill development and transfer. She varied her modeling events across different teaching contexts that considered teaching time and instructional materials. Repeated modeling allowed pre-service teachers to be able to identify and describe new reading comprehension strategies and incorporate them into their own instruction. Repeated modeling of classroom management techniques have also been accredited to improvements in preservice teacher retention and strategy implementation. Christofferson and Sullivan (2015) surveyed 157 preservice teachers throughout the United States to identify what training sources they attributed to their knowledge and skill development. In addition to their
coursework and hands-on training, respondents accredited repeated modeling and coaching as being most influential in their development of classroom management skills.

**Video modeling.** Symbolic modeling to promote task learning and self-efficacy building is relatively recent and more prevalent in the behavioral intervention literature primarily concerned with children with autism spectrum disorder. This literature examines the relevance of Video Based Interventions (VBI) for observational learning and self-efficacy promotion (Spencer, Evmenova, Boon, & Hayes-Harris, 2014). While this research base does not specifically examine the impact of video model on teacher education and self-efficacy development, it does offer useful guidance for utilizing terminology, which distinguishes types of VBI and levels of scaffolding assistance (Rayner, Denhom, & Sigafoos, 2009).

Schuman and Relihan (1990) assert that model observation is effective in teacher education programs for teaching whole language based pedagogical techniques, but only examine video modeling in terms of providing performance feedback and do not address the issue of teacher self-efficacy development resulting from live or video model observation. Reinke, Stormont, Webster-Straton, Newcomer, & Herman (2012) report on the effectiveness of a classroom management program in teacher professional development and make claims that the program builds teacher self-efficacy. Although video modeling is used as a core method to demonstrate the implementation of teaching practices, teachers also observed live model demonstrations, received tangible rewards and verbal encouragement from coaches, and reflected on their mastery experiences. These diverse self-efficacy information sources were not studied separately, nor was there any mention of instrumentation to actually measure teacher self-efficacy change. Studies such as these illustrate the sparse attention in the literature given to the study of video modeling to affect teacher self-efficacy development.
Although video modeling has not been specifically used to develop teacher self-efficacy, video observation has become an important component of teacher education and teacher professional development programs (Gaudin & Chaliès, 2015; Llinares & Valls, 2010; Santagata, 2009; Santagata, Gallimore, & Stigler, 2005). Observation of teaching models in videos can inform novices of the complexity and subtlety of realistic classroom situations (Brophy, 2004; Goldman, Pea, Barron, & Denny, 2014). In pre-service teacher education, videos can be an efficient means for illustrating how pedagogical theories may be put into practice to enhance student learning (Darling-Hammond & Bransford, 2005; Gomez, Sherin, Griesedorn, & Finn, 2008).

When video modeling is examined in the context of teacher education, its relevance to diverse pedagogical approaches to affect teacher learning are typically explored (Grossman, 2005). The use of videos in teacher education has been used as an effective tool to assist in teacher knowledge and skill development (Darling-Hammond, 2006a). Blomberg, Renkl, Sherin, Borko, & Seidel (2013) argue that research-based heuristics be used when teacher education programs make decisions on how to effectively use videos for pre-service teacher instruction. Their model specifies five cyclical steps to create a well-conceived learning environment which effectively employs video-based instruction: 1. Identify learning goals, 2. Align instructional approaches and learning activities with selected learning goals, 3. Carefully select learning videos to promote specified teacher learning, 4. Address limitations of selected videos, and 5. Align assessment practices with learning goals in order maximize the efficacy of the video-based learning environment and the engagement of pre-service teachers.

The literature base examining the use of videos in teacher education emphasizes that videos, or any form of multimedia, are neutral in nature and are only a means of delivering the
learning content, not the content itself (Brophy, 2004; Clark, 1994; Seago, 2004). Learning from models in videos can be differentially affected by the learning strategy in which they are situated (van Es, 2009). The distinct pedagogical approaches employed by the teacher education program may affect different aspects of novice teacher learning such as the depth and duration of their reflective processing after observation of teaching skills modeled in videos (Blomberg, Sherin, Renkl, Glogger, & Seidel, 2013).

Video modeling has also been studied for its capacity as a reflective instrument when teachers observe videos made of their own teaching behaviors or from other teachers’ classroom practice (Johnson & Golombek, 2016). Video observation allows novice teachers to experience a wider array of pedagogical practices, and to be able to replay modeled behaviors than would be available to them if they were to observe live teaching models (Bayram, 2012). Novice teachers have more opportunities to make sense of classroom practices, and to form important connections between their theoretical knowledge and its application in classrooms (Borko, Jacobs, Eiteljorg, & Pittman, 2008). Since videos are experienced first-hand, they may represent more compelling evidence on the effectiveness of recommended teaching practices than expository text or reported statistical information (Miller & Zhou, 2007). When novice teachers observe teaching models in videos they engage in more reflection and are able to provide a more extensive analysis of classroom interactions and contexts (Santagata, 2007; Stockero, 2008). Increased levels of sophistication in attention while viewing video models have been linked to more sophisticated analyses of targeted teaching practices (Barnhart & van Es, 2015).

Learning from video modeling and productive reflection can be improved when novice teachers are provided with scaffolding prompts that focuses their attention on relevant elements of the demonstrated teaching skill and salient aspects of the teaching context (Blomberg et al.,
Instructional prompting can counter novice teachers’ difficulties in recognizing and applying previously learned strategies (Flavell, 1979; Pintrich, 2002). From a cognitive viewpoint, direct instruction with a large amount of scaffolding is critical for novice teachers to reduce extraneous cognitive load (Sweller, 1988) and permit deeper expert-like reflection (Blomberg et al., 2013). Reflective processes may be enhanced when novice teachers are provided with semi-structured forms containing factual and/or reflective prompts. This process of guided reflection after observing teaching skills demonstrated on videos have been demonstrated to improve novice teacher involvement and quality of their reflective analysis (O’Sullivan, 2002). Scaffolding may be presented in a variety of formats such as written prompts (O’Sullivan, 2002; Seago, Mumme, & Branca, 2004), multimedia tools (van Es & Sherrin, 2002), or social interactions (van Es, 2009; van Es & Sherrin, 2008).

Social persuasion. Social persuasion is another source of information for self-efficacy appraisal. Social persuasion is the experience of the verbal judgments of others for one’s task performance or expected task performance (Pajares, 2002a). Positive social persuasion may promote increased effort and task persistence in individuals if they are convinced they have the capacity to master tasks. These behavioral changes are likely to promote skill development and related sense of self-efficacy (Bandura, 1994). However, positive persuasion is fleeting and can be easily undermined by negative persuasion or more powerful contradictory sources of self-efficacy appraisal (Bandura, 1986, 1994; Pajares, 2002a). Tschannen-Moran and Woolfolk Hoy (2007) determined that the interpersonal support of administrators, colleagues, parents, and members of the community, appeared to be more related to novice teachers’ self-efficacy than to those of career teachers. They concluded that career teachers relied on other sources to inform their self-efficacy and that social persuasion played a smaller role for teachers as support systems.
were withdrawn and they began to teach in their classrooms, socially isolated from other teachers. Johnson (2004) illustrates this point in her highly detailed descriptive analysis of social and structural supports that are conducive to retaining new teachers beyond their early years of professional employment.

**Physiological and affective feedback.** Physiological and affective states also appear to play a lesser role in determining teacher efficacy. Individuals rely somewhat on their physical and emotional reactions related to their performance when assessing their self-efficacy. The magnitude of their physical and emotional reactions are not relevant to their efficacy appraisals. Their perception and interpretation of these states determine the relative value for informing an individual’s efficacy. Researchers have largely discounted the impact of this source on self-efficacy determinations with the exception of studies of teacher stress and burnout (Brouwers & Tomic, 2000; Schwarzer & Hallum, 2008). Increases in teacher stress are associated with decreases in teacher self-efficacy and may be predictive of teacher burnout. Tschannen-Moran and her colleagues (1998) agree that physiological and affective states may contribute to informing teaching self-efficacy at moderate levels, but doubted that teachers may attend fully enough to these states for them to be of any consequence.

**Research exploring the sources of teacher self-efficacy.** Klassen and his colleagues (2011) review of teacher efficacy research from 1998 to 2009 found only 7 empirical articles concerning the sources of teacher self-efficacy. This echoes the findings of Tschannen-Moran and her colleagues in their review of the literature from 1986 to 1997. Only the research of Usher and Pajares (2009) was noted to have adequately addressed reliability and validity concerns.

Tschannen-Moran and Woolfolk Hoy (2007) determined that verbal persuasion, vicarious experiences, and physiological/emotional arousal sources of information for teacher efficacy
appraisals were more prominent for preservice teachers who had little significant mastery experiences. Experienced teachers, who had considerably more mastery experiences than novices, were less impacted by the presence of supporting contextual factors (i.e., school leadership, collegial support, and the availability of resources) for assessing their teacher efficacy. The four sources of information for self-efficacy appraisal appear to effect teachers differentially depending on contextual factors and individual interpretation differences. However, none of the studies in Klassen’s review examined the impact of vicarious observation isolated from other sources of teacher self-efficacy (mastery experiences, persuasion, and bio-affective feedback).

**Malleability of Self-efficacy**

Just as the individuals develop their self-efficacy by interpreting information from the four theorized sources mentioned above, their self-efficacy appraisals may similarly be adjusted when they encounter new information from these same sources (Bandura, 1977b; Pajares, 2002a). Feedback that is credible, explicit, and compelling may alter preexisting self-efficacy depending on the strength of an individual’s perceived self-efficacy and the focus of their reflective process (Bandura, 1997). An individual’s appraisal of their self-efficacy may change over time when new intervening information and experiences are considered. “The degree of change that might be expected in self-efficacy is partly a function of the variability and locus of its determinants” (Gist & Mitchell, 1992, p. 200). Teachers who participate in collaborative experiences with their fellow educators are exposed to more sources of vicarious experience, social persuasion, and performance feedback that are likely to influence their self-efficacy appraisals depending on how compelling and credible the model and feedback are.
An individual’s self-efficacy may also function as a self-fulfilling prophecy because of the effect it has on the amount of effort and perseverance they display when engaging in challenging tasks (Bandura, 1986; Pajares, 2002a). Higher levels of self-efficacy are likely to contribute to increased effort and perseverance in task execution, which may lead to improved performance. An individual’s self-efficacy appraisals are likely to increase when they reflect on their improved mastery experiences. Similarly, lower levels of self-efficacy are likely to have negative effects on an individual’s amount of effort and persistence, which, in turn may lead to task failure or lowered performance resulting in decreased self-efficacy. Eden and Aviram (1993) demonstrated this effect in their study of a 2 ½ week behavioral modeling workshop. They found that participants with initially lower self-efficacy were able to increase their effort in searching for employment after observing positively modeled job search behaviors and practicing these behaviors with constructive feedback from their instructors. Results demonstrated that preexisting self-efficacy was positively correlated with participants’ job search behavior, and increasing self-efficacy led to an intensification of job search behaviors. Subsequently, when participants increased their job search efforts, they were more likely to obtain employment. The authors concluded that augmenting participants’ low self-efficacy allowed them to break through their negative cycle of low effort and failure in their search for employment.

How individuals characterize their abilities are also important to the malleability of their self-efficacy. Wood and Bandura (1989) demonstrated that the manner in which individuals interpret the source of their ability can affect the stability of their self-efficacy when they encounter disconfirmatory evidence. Individuals who construed their ability as an acquired skill demonstrated a highly resilient sense of efficacy to perform challenging new tasks, while
individuals who conceptualized their ability as being indicative of an inherited aptitude sharply decreased the strength of their self-efficacy appraisals after they encountered difficulties in their task performance.

In addition to reinterpreting new information on one’s capacity to adequately perform a known task, self-efficacy may also change when the parameters or context of a task changes. Challenges to teacher self-efficacy may come from sudden changes in the context of the teaching environment. Tschannen-Moran and her colleagues (1998) suggested that teachers’ judgments of self-efficacy may be affected by their analysis of their competence when the perceived demands for a teaching task are altered. Rapid changes in technology in the workplace demand higher problem-solving skills and resiliency in one’s self-efficacy to cope with changes in vocational activities (Bandura, 1994). Teacher self-efficacy often increases during teacher preparation but falls when teachers experience authentic teaching demands in their first teaching placements (Woolfolk Hoy, 2000). Teachers also experience challenges to the self-efficacy when asked to implement new teaching innovations (Bruce & Ross, 2008; Ghaith & Yaghi, 1997; Gregoire, 2003; Guskey, 1988; Wheatley, 2005). There is some evidence that challenges to teaching efficacy represented by contextual changes may also be insulated by contextual supports, especially among novice teachers who are unable to rely on many mastery experiences to provide their teaching efficacy with a sense of resiliency (Knoblauch & Woolfolk Hoy, 2007).

**Self-efficacy’s Influence on Human Functioning**

Self-efficacy can improve human functioning in a variety of ways. Self-efficacy influence the way individuals think, feel, motivate themselves, and behave. Perception of self-efficacy acts as a regulatory mechanism influencing behavior through four major processes: cognitive, motivational, affective, and selection processes (Bandura, 1992, 1993; Pajares, 2002a). The self-
regulatory nature of self-efficacy can be understood within the conceptual framework for how individuals implement human agency within social cognitive theory (Bandura, 1993). Bandura (1997) asserted that self-efficacy plays a critical role in human functioning because “people’s level of motivation, affective states, and actions are based more on what they believe than on what is objectively true” (p.2). An individual’s self-efficacy is a strong determinant for what he or she actually does with the knowledge and skills he or she possess (Pajares, 2002a).

Most human behavior is purposive and regulated by anticipatory thinking that encompasses an individual’s cognitive awareness of their goals. Plans to achieve goals are structured by an individual’s cognitions and mediated by their self-efficacy as they construct and visualize anticipatory scenarios (Bandura, 1993). Individuals who have higher efficacy for implementing behaviors related to their goals will create scenarios that are helpful to their goal achievement, while individuals who have a lower sense of efficacy will visualize unsuccessful scenarios and face obstacles to successful performance created by their self-doubts. “Personal accomplishments require not only skills but self-efficacy to use them well. Hence, a person with the same knowledge and skills may perform poorly, adequately, or extraordinarily depending on fluctuations in self-efficacy thinking” (Bandura, 1993, p. 119). Individuals’ evaluation of their abilities is regulated by self-efficacy through its effects on their thought processes and achievement. Pintrich and De Groot (1990) suggested that self-efficacy mediates levels of cognitive engagement and that increased efficacy may lead to more frequent use of cognitive strategies that will facilitate individual performance. Wood and Bandura (1989) conducted a study where participants were asked to take on the role of a manager in a simulation where they were required to decide how to appropriately match employees to particular tasks, and to discover and apply rules of management to achieve higher levels of organizational performance.
Participants were either told that their capacity to execute managerial tasks were based either on an inherent intellectual ability or on an acquirable intellectual skill. Participant conceptions of ability and their performances differed by group. Participants who viewed ability as an inherited trait had decreased efficacy as they encountered challenges. They demonstrated difficulties in thinking analytically about problems and had decreasing achievement over time, eventually lowering their organizational goals. Conversely, participants who viewed ability as a temporary state had stronger and more consistent levels of efficacy. They were better able to maintain their use of analytic problem solving strategies and demonstrated higher levels of performance with increased constancy in setting challenging organizational goals. This study illustrated the effects of individuals’ cognitive evaluations of the stability feature of their causal attributions (i.e., inherited vs learned) on their self-efficacy appraisals and related behaviors.

Motivation processes are also regulated by self-efficacy. Bandura (1994) made connections between his theory of self-efficacy and other cognitive theories of motivation: attribution theory, expectancy-value theory, and goal theory. He suggested that self-efficacy contributed to our understanding of individuals’ causal attributions, outcome expectancies, and cognized goals. Individuals consider their self-efficacy when engaging in these types of cognitive motivation processes. When individuals make attributions for their performance failures, those with high self-efficacy are likely to consider their amount of effort as being inadequate; while those with low self-efficacy are more likely to believe it was due to their inherent lack of ability. Individuals’ outcome expectancies are also mediated by their self-efficacy. As mentioned above, an individual’s self-efficacy have causal influence over their outcome expectancies. People don’t pursue many courses of action because they believe they lack the capacity to perform at adequate levels for successful task performance (Bandura, 1986,
Individuals are more likely to engage in tasks they feel efficacious for performing successfully, they persist longer when confronted by obstacles, expend greater amounts of effort, and see difficulties as challenges to be mastered rather than threats to be avoided (Bandura, 1994; Pajares & Schunk, 2001). People with high self-efficacy for given tasks set more challenging goals and experience a stronger sense of commitment toward their achievement (Bandura, 1977a, 1977b, 1986, 1997; Pajares, 1997, 2002a). There is little incentive for individuals to initiate and persist in tasks they believe will not lead to desired outcomes.

Perceived self-efficacy also influences individuals’ affective states. Their perceived ability to cope with demanding circumstances impacts the amount of stress, anxiety, and depression they experience (Bandura, 1994). Individuals face challenges to their functioning and sense of well-being if they perceive that they are unable to control their thoughts which cause them distress. Teachers experiencing low self-efficacy for coping with challenging situations face higher levels of stress, job dissatisfaction and burnout (Brouwers & Tomic, 2000; Fives, Hamman, & Olivarez, 2007; Klassen & Chiu, 2010, Schwarzer & Hallum, 2008; Skaalvik, & Skaalvik, 2007). Conversely, higher levels of teacher efficacy were related to lower stress levels among teachers and improved stress management ability (Johnson, 2004; Parkay, Greenwood, Olejnik, & Proller, 1988).

Self-efficacy also influences an individual’s selection processes. Their level of efficacy affects not only their choice of activities, but also the amount of effort they exert and the duration of their persistence while performing behaviors related to the activity (Bandura, 1977b). Feelings of competency and confidence effect what tasks individuals choose to engage in. People are more apt to select tasks and activities that they have a higher sense of efficacy for performing competently than for those that they don’t (Bandura, 1977b, 1986, 1997; Pajares, 2002a; Schunk,
 Individuals avoid tasks and circumstances they believe present intractable challenges to their adaptive capacities (Bandura, 1994). Individuals have more incentive for engaging in behavior they believe that they can competently perform and consequently lead to desirable outcomes. However, choice is not a discrete, one-time event since “the social influences operating in selected environments continue to promote certain competencies, values, and interests long after the efficacy decisional determinant has rendered its inaugurating effect” (Bandura, 1994, p. 6). Most studies of self-efficacy’s influence over selection processes focus on career choice or professional growth (see for example Hacket, 1995; Lent & Hackett, 1987). As already mentioned, the ongoing choice to either remain in the profession or leave prematurely, due to the impact of teacher burnout effects, is strongly determined by the strength and resiliency of teacher efficacy. Teachers must also respond to the daily educational needs of their students and there are strong suggestions in the literature that teacher efficacy is related to important instructional decisions linked to their use of instructional time, classroom management strategies, the types of questions they pose to students, and the manner in which they engage their students (Gibson & Dembo, 1984; Klassen et al., 2009; Ross & Bruce, 2007; Skalofske, Michayluk, & Randhawa, 1988; Woolfolk, Rosoff, & Hoy, 1990).

**Conceptualizing and Measuring Teacher Self-efficacy**

Teacher self-efficacy refers to an educator’s personal beliefs about his or her capabilities to help their students learn (Schunk, 2011; Woolfolk Hoy & Burke Spero, 2005). Teacher self-efficacy was first conceptualized in terms of internal and external loci of control for affecting student performance (Rotter, 1966). The Rand Report for the Los Angeles Unified School District had a 2-item measure (Armor et al., 1976) asking teachers to assess their beliefs about the extent to which student outcomes were within or outside of their control. Shortly after, more
lengthy instruments were developed by researchers who had concerns about the reliability of RAND’s two-item measure (Tschannen-Moran & Woolfolk Hoy, 2001). The Responsibility for Student Achievement (RSA) instrument (Guskey, 1981), Teacher Locus of Control (TLC) instrument (Rose & Medway, 1981), and the Webb scale (Ashton, Buhr, & Crocker, 1984) were all introduced within a short time of each other and conceptualized teacher efficacy as a teacher’s ability to control student outcomes. Although these measures considered efficacy expectancy or “the individual’s conviction that he or she can orchestrate the necessary actions to perform a given task” (Tschannen-Moran & Woolfolk Hoy, 2001, p. 787), their use of outcome expectancy was unproductive since it added little to their predictive power of their instrument.

Notwithstanding minor debate in the literature questioning the nonreciprocal relationship between efficacy expectancy and outcome expectancy (Williams, 2010), current and more widely accepted measures adhere to Bandura’s (1997, 2006a) theoretical and practical applications.

Conceptualization of teacher self-efficacy within social cognitive theory (Bandura, 1977b, Schunk, 2011) proved to be a much more productive line of research. Gibson and Dembo (1984) attempted to improve on the reliability and validity of the Rand two-item measure by constructing a 30-item instrument - the Teacher Efficacy Scale (TES). They also attempted to apply Bandura’s 2-dimensional construct of expectancy to their instrument design within social-cognitive theory’s (Bandura, 1986) framework (i.e., self-efficacy and outcome expectancy). Gibson and Dembo attempted to equate the locus of control elements of the Rand study with the self-efficacy and outcome expectancy elements of Bandura’s theory. They defined one dimension of their teacher self-efficacy scale as general teaching efficacy (GTE) and the other dimension as personal teaching efficacy (PTE). GTE was thought to correspond to the first Rand
item measuring external locus of control and Bandura’s outcome expectancy construct. Similarly, PTE was thought to correspond to the second Rand item measuring internal locus of control and Bandura’s self-efficacy construct. However, later factor analysis (Woolfolk & Hoy, 1990) of Gibson and Dembo’s instrument indicated that GTE was not consistent with Bandura’s outcome expectancy construct, but more reflective of a general belief about the ability of teaching to positively impact difficult children. Although GTE diverges from other measures of teacher self-efficacy (Woolfolk Hoy & Burke Spero, 2005), it’s value lies in its ability to represent teachers’ general beliefs about the ability of teaching to influence students with challenging learning circumstances and may indicate teachers conservative or liberal beliefs about education (Tschanne-Moran et al., 1998). Irrespective of findings contradicting the TES’ original conceptualization of GTE and item factor loadings (Soodak & Podell, 1993), the Gibson and Dembo’s instrument has subsequently become the most used instrument in the study of teacher efficacy and has also served as a template for the development of other similar subject matter specific instruments (Henson, 2001; Ross, 1994). However, careful attention should be given the literature addressing the validity the TES’ two factors during the interpretation of scores from this instrument.

Subsequent attempts to measure teaching efficacy have shifted away from broad conceptualizations of teacher functioning toward more specific multidimensional areas that appear to be most relevant when teachers consider their teaching performance. Bandura attempted to address this concern when he introduced his own 30-item instrument (Bandura, n.d.). This unpublished instrument attempts to provide a multidimensional measure of teachers’ self-efficacy while still maintaining generalized sub-scales that could be applied within different teaching domains. The seven sub-scales of this instrument include: influence on decision
making, influence on school resources, instructional efficacy, disciplinary efficacy, enlisting parental involvement, enlisting community involvement, and creating a positive school climate. Tschannen-Moran and Wolfolk Hoy (2001) developed the Teachers’ Sense of Efficacy Scale (TSES) by polling teachers on elements of Bandura’s scale plus additional variables they thought important to teacher functioning. The TSES is a modification of Bandura’s multifaceted unpublished instrument but is thought to more accurately represent the nature of tasks normally encountered by teachers. The instrument measures teacher self-efficacy along three subscales: teacher efficacy for instructional strategies, teacher efficacy for classroom management, and teacher efficacy for student engagement. Recently, Skaalvik and Skaalvik (2007) analyzed teacher role expectations in Norwegian schools to develop the Norwegian Teacher Self-efficacy Scale (NTSES). They attempted to improve on the TSES by constructing items more closely with Bandura’s (1997) assertion that items measuring self-efficacy should contain clear obstacles to surmount. Otherwise, self-efficacy for items would not be meaningful and items would fail to discriminate between respondents. Their instrument measures teachers’ self-efficacy along six subscales: instruction, adapting education to individual students’ needs, motivating students, keeping discipline, cooperating with colleagues and parents, and coping with changes and challenges.

Teacher Self-efficacy Development

Teacher efficacy development has been conceptualized as a cyclical process (Figure 1). Teachers make appraisals of their capabilities by cognitively processing various sources of efficacy information to develop a sense of their teaching competence. This discernment is then integrated with their assessment of the difficulty of the teaching task, understanding of their specific teaching context, and their calculation of their personal teaching competence. During
this analysis, teachers are likely to rely on the causal attributions they make for their task success or failure (Bandura, 1994). Teachers who more consistently make effort attributions see outcomes as within their control if they try harder and persist longer. Increased effort and persistence will likely improve their teaching performance and subsequently their related self-efficacy for teaching at a competent level to achieve desired positive outcomes (Bandura, 1997; Pajares, 2002). Their adjusted teaching efficacy then leads to associated outcomes (e.g. – possibly increased task persistence or goal reassessment). These behavioral and belief outcomes consequently affect the quality and magnitude of their teaching performance creating the potential for a possible mastery experience or task failure experience which in turn generates
new evidence to inform their future teaching self-efficacy appraisals. For pre-service teachers, these performances have less relevance because these so-called mastery experiences are sheltered and usually performed under the watchful eye and guidance of an expert teacher educator. These experiences also do not generalize very well between disparate teaching contexts for pre-service teachers entering their new placements (Brandt, 2006; Faez & Valeo, 2012; Favre & Knight, 2016, Tschannen-Moran et al., 1998).

**Longitudinal studies on teacher self-efficacy development.** Although Tschannen-Moran and her colleagues (1998) called for more longitudinal studies to better determine how teacher self-efficacy develops, only 19 studies were later conducted which employed multiple measures over phased time periods (Klassen et al., 2011). However, these studies either simply examined longitudinal post-treatment effects or phased periods over a treatment period that included exposure to multiple sources of self-efficacy information. Most studies examined longitudinal effects of teacher self-efficacy on their rate of professional burnout or job stress. Woolfolk Hoy and Burke Spero (2005) did examine teacher self-efficacy development in three phases: their first quarter of teacher training, at the end of their teacher training and student-teaching, and at the end of their induction year. Teachers’ perceptions of self-efficacy increased between the first two phases, but declined during actual teaching. This decline was attributed to perceptions of support available within their teaching placements.

Zee and Koomen’s (2016) recent review of the last 40 years of teacher self-efficacy research on important educational processes and outcomes found a similar emphasis on teacher self-efficacy as an independent variable. They noted two studies that found that teacher experience had positive effects on teacher self-efficacy, but this beneficial effect decreased over
time. They emphasized that findings such as these point to the importance for conducting longitudinal investigations of teacher self-efficacy development.

Interventions aimed at increasing teacher self-efficacy should attempt to manipulate variables affecting teachers’ interpretation of self-efficacy source information or their perceptions of the demands of targeted teaching tasks (Pajares, 1996). Wheatley (2005) suggested that frequent measurement of teachers’ self-efficacy needs to be built in to a longitudinal study of their development in teacher education programs. Phased studies with simple pre- and post-measures are unable to adequately isolate the variables affecting teacher self-efficacy.

**Teacher Self-efficacy’s Impact on Educational Outcomes**

The importance of teacher efficacy for obtaining desirable educational outcomes is well documented within the literature (Ross, 1998; Tschannen-Moran et al., 1998; Zee & Koomen, 2016). There is convincing evidence from research demonstrating an association between teachers’ self-efficacy, and important school processes and outcomes (Tschannen-Moran & Wolfolk Hoy, 2001). Teacher self-efficacy has been related to their instructional practices and to student outcomes such as student achievement (Armor et al., 1976; Ashton & Webb, 1986; Bandura, 1993; Goddard, 2001; Goodard, Hoy, & Woolfolk Hoy, 2004; Ross, 1992, 1998), student beliefs in their own self-efficacy (Anderson, Greene, & Loewen, 1988; Knoblauch & Woolfolk Hoy, 2007), student motivation, (Midgley, Feldlaufer, & Eccles, 1989; Tschannen-Moran & Wolfolk Hoy, 2007), more positive student attitudes toward school and teachers (Woolfolk, Rosoff, & Hoy, 1990) and student emotional growth (Tschannen-Moran et al., 1998).

At the classroom and school levels, teacher efficacy has been related to successful implementation of instructional innovation and educational reform teaching practices (Allinder,
1994; de Mesquita & Drake, 1994; Duran, Ballone-Duran, Haney, & Beltyukova, 2009; Ghaith & Yaghi, 1997; Guskey, 1988; Ross, 1998; Sharma, Loreman, & Forlin, 2011; Smylie, 1988, Stein & Wang, 1988, Tschannen-Moran & McMaster, 2009). Student outcomes are also critically important factors assessed by institutions judging teacher and school effectiveness, and appear to have a strong reciprocal relationship with teacher efficacy (Hoy & Woolfolk, 1993). Positive student outcomes can inform teacher self-efficacy, which in turn can contribute to further improvements to student outcomes. There also appears to be a similar reciprocal relationship between teacher self-efficacy and student motivation (Brophy, 1983). Students who respond enthusiastically to a teacher’s instruction may encourage a teacher to plan more interesting or exciting lessons, which in turn may positively impact student levels of motivation in the classroom.

Teaching efficacy has also been linked with different positive and negative teacher traits and outcomes. Bandura’s hypothesis that increased self-efficacy would lead individuals to expend more effort on difficult tasks and to persist longer has been supported in the research on teachers in their classrooms (Ashton & Webb, 1986; Gibson & Dembo, 1984; Tschannen-Moran et al., 1998, Wolfolk & Hoy, 1990). Task persistence and expenditures of effort that lead to a sense of mastery can, in turn, inform self-efficacy (Ames, 1992). Ashton and Webb (1986) also found that self-efficacious teachers were more likely to maintain a positive classroom environment, support students’ ideas and assist students with their needs. Positive correlations have also been noted between teachers’ constructivist oriented teaching beliefs and their teacher efficacy about student engagement; and between traditional teacher beliefs and teacher efficacy about class management, instruction and overall self-efficacy (Gürbüztürk & Şad, 2009). Teachers with a strong sense of efficacy demonstrate more effort in teacher planning and
encouraging students during their interactions. They are also more able to adopt a learner-centered approach to their instruction (Gürbüztürk & Şad, 2009; Magno, 2007). Teachers who possess a strong sense of efficacy also show more commitment to their work (Chan, Lau, Nie, Lim, & Hogan, 2008). Johnson (2004) found that teachers’ sense of efficacy “infused their accounts of job satisfaction and career plans” (p. 81) with higher levels of efficacy serving to promote a sense of positive functioning and affirmation that they were well suited for their profession. Weak levels of teaching efficacy have been related to significant negative outcomes such as teacher stress, and burnout (Brouwers & Tomic, 2000; Fives, Hamman, & Olivarez, 2007; Klassen & Chiu, 2010; Schwarzer & Hallum, 2008; Skaalvik, & Skaalvik, 2007).

Determining Teacher Self-efficacy

Assessment of self-efficacy should address the specific questions held by researchers and the types of judgments that individuals make when considering their capacity to perform a given task. Bandura (1986) asserted that researchers must consider types of judgments individuals make as they consider the capacity to perform specific behavioral tasks. He believed that reasonably precise judgments of ability that corresponded to a specific outcome were needed to optimize an assessment’s predictive utility. Judgments of self-efficacy rely on a calculation of performance expectations and contextual framing. Individuals reference their goal directed behavior in relation to task-specific and situation-specific contextual factors (Pajares, 1997).

Bandura (1997) stated that “self-efficacy should be measured in terms of particularized judgments of capability that may vary across realms of activity, different levels of task demands within a given activity domain, and under different situational circumstances” (p. 6). Pajares (1996) review of assessment of self-efficacy in academic settings supported Bandura’s assertion. Pajares found that “particularized measures of self-efficacy that correspond to the critical tasks
with which they are compared surpass global measures in the explanation and prediction of related outcomes” (p. 543). Particularized measurement of teachers’ self-efficacy require an understanding of what specific judgments they make for their ability to perform specific behaviors related to identified areas of functioning. Pajares believed that the core issue in self-efficacy measurement centered around which types of questions are asked and how they are related to the beliefs of the individual. As already mentioned, critical performance expectations for adequate teacher functioning may be identified by summarizing trends from teacher interviews and polling (Tschannen-Moran & Wolfolk Hoy, 2001) or from an examination of formalized role expectations (Skaalvik and Skaalvik, 2007). Appropriate measurement of self-efficacy considers variance across three dimensions: strength, magnitude, and generality (Bandura, 1977b, 1986; Pajares, 1997).

**Strength.** The strength of an individual’s self-efficacy denotes the length of time the individual continues to believe that he or she can succeed in performing a behavior at an expected level of adequate competency (Bandura, 1977b, 1986). As previously mentioned, individuals with strong self-efficacy persist longer at tasks as they experience challenges, setbacks, and failure. Individuals with weak self-efficacy give up more readily when they experience difficulties and failure.

**Magnitude.** The magnitude of self-efficacy refers to the degree of confidence that an individual believes he or she can successfully perform a behavior at a designated level and is strongly influenced by his or her perception of task difficulty (Bandura, 1977; 1986). Bandura (2006a) stated that the magnitude of an individual’s self-efficacy appraisal should be determined by asking the individual to assess his or her level of confidence for current operative capabilities for a specific activity. He illustrates this point with his suggestion for providing physical tasks
for children to familiarize themselves with the increasing level of confidence values. Children can be asked how confident they believe they can lift progressively heavier weights or jump increasingly longer distances. Bandura advises that scales should be anchored by descriptors indicating progressively stronger levels of confidence and that researchers should take care in determining reasonable cutoff values of efficacy magnitude. Low cutoff values may not be able to separate individuals with relative low self-efficacy from individuals who have complete confidence. Items should also reflect prototypical tasks encountered by individuals in their area of expected functioning (Pajares, 2007).

Determining the magnitude of self-efficacy can also be facilitated by item analysis during scale construction. Bandura (2006a) instructs that all items should be pretested with ambiguous items being rewritten or discarded. Items where individuals indicate the same response should be discarded. Items where most individuals indicate maximum magnitude of efficacy should be revised to provide sufficient difficulty or challenge to task completion so that variance among respondents may be distinguished. Capturing the magnitude of individuals’ self-efficacy requires that items be anchored by increasingly higher levels of expressed confidence across various gradations of challenge (Bandura 1997, 2006a).

**Generality.** Self-efficacy is more predictive of performances that most closely correspond to specific areas of activity and situational circumstances (Pajares, 1997). Bandura (1997) found that many instruments measuring self-efficacy were often constructed too broadly. He suggested finding an appropriate balance between generality and specificity for self-efficacy measures. If an item is too general, teachers may not be able to properly conceptualize the task category to provide an accurate assessment of their efficacy. If an item is too specific, then any
meaningful generalization of teacher efficacy beyond the exact context of the item would be difficult to justify.

Considerations for the level of item specificity are rarely reported in validations of teacher efficacy instrumentation. Bandura has provided some guidance to researchers for striking the proper balance between generality and specificity. He advised researchers that “the optimal level of generality at which self-efficacy is assessed varies depending on what one seeks to predict and the degree of foreknowledge of the situational demands” (1986, p. 49). Bandura (2006) also warned against using global measures of perceived self-efficacy because items on such assessments reflect a great deal of ambiguity for what is being probed and much of the contextual references are obscured. Because most items have little relevance to the domain of functioning they purport to assess, such general measures of teacher efficacy have limited explanatory and predictive value.

The connection of self-efficacy appraisal to the appropriate level of task is critical to the construct validity of the interpretation of scores for an assessment instrument (Gist & Mitchell, 1992). What cognitive considerations do teachers make when they are asked to appraise their capacity to perform a task at a specific level? Pajares (1996) expanded on this concern for striking an appropriate balance between specificity and generality of item focus:

Omnibus tests that aim to assess general self-efficacy provide global scores that decontextualize the self-efficacy-behavior correspondence and transform self-efficacy into a generalized personality trait rather than the context-specific judgments Bandura suggests they are. Generalized self-efficacy instruments assess people's general confidence that they can succeed at tasks and in situations without specifying what these tasks or situations are. Even domain-specific omnibus measures are problematic if
composite multiscale scores drawn from differing subsections of the domain are used. (p. 547)

Pajares (1996) also warned researchers about going to the other extreme when constructing efficacy measures when he observed that “specificity and precision are often purchased at the expense of external validity and practical relevance” (p. 561).

Bong (2006) likewise recommended determining the appropriate level of specificity when constructing self-efficacy assessment instrumentation. She suggests that finding the most suitable balance in item construction requires a careful analysis of the targeted task performance considered for prediction.

If the goal is to predict students’ levels of performance on some defined academic task, researchers must analyze the types of skills, knowledge, and potential constraints on performance involved in successful accomplishment of the task before creating a self-efficacy scale. (p. 296)

Unfortunately, finding the right level of performance for teacher efficacy measurement has been elusive or beyond the considerations of many researchers. “Perhaps the greatest challenge has to do with finding the appropriate level of specificity for measurement” (Tschannen-Moran et al., 1998, p. 219).

Teachers are required to perform innumerable tasks in many areas of their professional functioning. When assessing teacher self-efficacy, it is critical that researchers consider teachers’ particularized beliefs for their capacity to adequately perform specific tasks across appropriately identified strength, magnitude, and generality parameters. Increased attention to the theoretical conceptualizations guiding the language and focus of survey items is critically important to establishing evidence for claims of assessment validity.
Summary

The review of the literature indicates that researchers have generally not employed productive methods to measure self-efficacy development for novice teachers. Consequently, teacher education programs have very little guidance for how they can design programming to promote self-efficacy that has been consistently linked with many positive educational outcomes when novice teachers enter the workforce. Traditional procedures for measuring teacher self-efficacy do not adhere to theoretical notions of how to measure its development. If teacher self-efficacy is indeed malleable during early teacher education, then it is more likely to be strengthened by the social comparison that occurs during model observation rather than by infrequent and inauthentic mastery experiences. Small incremental changes may occur that necessitate frequent and repeated observations. Great care needs to be taken with designing self-efficacy questions that target the correct level of specificity for relevant teaching behaviors. Measurement should also be responsive to temporal concerns as well to avoid survey respondents experiencing other events that would affect the accurate assessment of any single treatment on specific self-efficacy.

The literature on teacher self-efficacy does indicate it can be an elusive concept to define and measure, especially for its development and promotion within novice teachers. However, if we repeatedly measure changes in teacher self-efficacy (as novices observe modeled teaching behaviors) directly after the self-efficacy informing event and address the correct level of specificity in assessment prompts, we may begin to observe self-efficacy development in real time as it is being affected by interventions that teacher educational programs would have some control over. Pilot studies were conducted to determine the viability of these procedures and validity of instrumentation necessary to measure teacher self-efficacy development in a
theoretically sound manner consistent with the review of the literature. These methods and results are reported in chapter 3.
Chapter 3 – Pilot Studies

Two pilot studies were conducted in order to develop and validate the materials and methods that are used in the dissertation study as well as to explore the viability of the study’s three research questions. The first pilot study explored participation rates, accessibility of materials, and the viability of research questions. Accordingly, the first pilot study examined the feasibility of the dissertation study’s three research questions:

1. How does observation of video modeling affect the magnitude of teachers’ micro-appraisals of their self-efficacy for performing the modeled teaching behaviors?

2. What is the relationship between teachers’ micro-appraisals for specific modeled teaching behaviors and their overall self-efficacy assessment of the larger sub-scale to which these modeled behaviors represent? (i.e., self-efficacy for performing instructional strategies)

3. Is there a relationship between causal attributions (i.e., effort, ability, luck, and task difficulty) and changes in the magnitude of teachers’ micro-appraisals of their self-efficacy for specific behaviors they observe in video modeling?

The second pilot study determined the validity of instrumentation to be used in the dissertation study for assessing participant micro-appraisals of their self-efficacy to perform the teaching skills they observe modeled in video presentations and to discriminate assessments of their self-efficacy from their assessments of their knowledge of these skills.

Pilot Study 1

The first pilot study was conducted to determine barriers to participation, accessibility of study materials and the viability of the initial three research questions. Participants (n=118) were recruited from a convenience sample of undergraduate students attending four separate introductory courses in educational psychology. Participants were randomly assigned to one of
three groups and were sent weekly links to their sessions on a Qualtrics survey platform where they were asked to view their assigned videos and afterwards assess their self-efficacy for performing the teaching behaviors they observed (micro appraisal). All participants could complete their sessions during each of the six weeks at a time of their convenience.

**Materials.** During each of the six weeks, the two treatment groups (Group A and Group B) and the control group (Group C) were presented with different video packages. Group A received four videos consisting of a direct instruction video, which explained the targeted pedagogical skill, and three modeling videos where the targeted teaching behavior was demonstrated in authentic contexts. Group B viewed the same three modeling videos that were presented to Group A, but was not presented with the direct instruction video. Group C was presented with only one video which discussed education issues but did not address any pedagogical techniques. Control group videos were of equivalent length to experimental group video presentations to create similar time-on-task conditions. The content of the video presentations are described below and links to each of the videos are provided in Appendix A:

**Week 1 videos:** The focus of this week’s pedagogical skill was on recasting. The direct instruction video (presented to Group A) defined recasting as “a form of feedback you can give students when they are answering questions”. The video provides several examples of language errors produced by English language learners (ELLs) that could be addressed by the teacher by revising the spoken error during the flow of their conversation with their students. The video describes several levels of possible teacher response variations that could scaffold teacher recast interventions based on the nature of the language error produced by the ELL student. The three modeling videos (presented to Groups A and B) consisted of teachers working in a variety of contexts where they corrected students’ spoken language errors using the recasting method. One
video had captioning so that language errors could also be visually identified by study participants. During this week, Group C viewed an unrelated video of equivalent length that focused on education policy research.

**Week 2 videos:** The focus of this week’s pedagogical skill was semantic gradients. The direct instruction video (presented to Group A) defined semantic gradients or gradable opposites by presenting two opposite meaning words, common to students, and putting them on far points along a visual word continuum. The video used “hot” and “cold” as an example and filled in points along the continuum with other known degree variations such as “scorching, warm, cool, and freezing” and a word unknown to students - “lukewarm”. The unknown word’s position along the continuum indicates its meaning to students. The video also described situations in which gradable opposites could be used and presented variations in visually displaying semantic gradient word clusters. The three modeling videos (presented to Groups A and B) demonstrated the semantic gradient technique for teaching students subtle word degree discriminations in a variety of contexts such as with adult ELLs, elementary students, and even two students who were using the technique to construct their own semantic gradient continuum using an outside reference. During this week, Group C viewed an unrelated video of equivalent length that focused on the unique system of education in the United States.

**Week 3 videos:** The focus of this week’s pedagogical skill was the finger correction technique. The direct instruction video (presented to Group A) defined finger correction as consisting of five steps to correct an ELL’s spoken language error: Step 1. A pause by the teacher to indicate the error and potential for possible student consideration, Step 2. Teacher asks the student to repeat what he/she spoke to see if the error was self-corrected, Step 3. The teacher “loads” the words onto his or her fingers by having the student repeat their sentence slowly and
the teacher raises each of his or her fingers word-by-word until the sentence is complete, Step 4. The teacher asks the student to repeat the sentence again with the teacher pointing to and pausing at the finger that is paired with the spoken error, Step 5. The teacher visually illustrates removing, extending, or adding a word or parts of a word that would form a correct sentence and asks the student to alter their sentence structure. Each of these steps provide increasingly scaffolded instruction and would only be necessary up to the step where the ELL was able to make the necessary correction. The three modeling videos (presented to Groups A and B) demonstrated the finger correction technique in a variety of teaching contexts including a foreign classroom, with the teacher also referring to visual aids on a whiteboard, and a point-of-view presentation with the teacher alone simulating an online learning context. During this week, Group C viewed an unrelated video of equivalent length that focused on redefining the national school curriculum to target important life skills necessary for students’ future success.

**Week 4 videos:** The focus of this week’s pedagogical skill was the board race technique. The direct instruction video (presented to Group A) defined board races as students practicing language recognition and production by forming two teams and running to the board to write their responses to the teacher’s grammar/vocabulary questions. Students take turns responding by running to the board to write their answer then running back to their team to handoff their markers to one of their teammates. The highest scoring team wins and the teacher reviews the accuracy of answers with the class and may also provide extension activities to further practice the targeted language skills. This activity can be used to practice and reinforce language instruction and a board race activity may be more engaging to students than learning language rules or completing deskwork. The three modeling videos (presented to Groups A and B) demonstrated the board race technique in varying adult education and elementary learning
contexts and covered topics such as comparative adjective construction, pronoun identification and use, and holiday vocabulary recall and production. During this week, Group C viewed an unrelated video of equivalent length that focused on obtaining a master’s degree in higher education administration at a small university.

**Week 5 videos:** The focus of this week’s pedagogical skill was using the task-based learning method for English language instruction. The direct instruction video (presented to Group A) defined the task-based learning method for learning English by “having students collaborate to solve a problem or complete an assigned task” while using necessary conversation skills, grammar, and vocabulary. The teacher provides the necessary language features (i.e., prior knowledge) to the students that they are to use to complete their assigned tasks. The teacher can elaborate on task requirements in a handout of instructions to increase student precision and provide a structure to adhere to so that students can maintain their focus on using targeted language features. The three modeling videos (presented to Groups A and B) demonstrated the task-based learning method for English language instruction with different task-based problem contexts such as creating a utopian society and making a map with location labels of banks, hospitals, etc.; going on a shopping trip with limited funds and trying to maximize group spending; and setting up a task-based learning activity and giving ELLs thorough activity instructions. Videos all showed adult learners in a variety of second language learning contexts. During this week, Group C viewed an unrelated video of equivalent length that focused on using data analytics to provide teachers with information that would be supportive of their administrative tasks and provide them with a comprehensive means to capture and analyze data.

**Week 6 videos:** The focus of this week’s pedagogical skill was promoting student language production through the use of the inside-outside circles method. The direct instruction
video (presented to Group A) defined using inside-outside circles as having two groups of students face each other in concentric circles standing or sitting at desks. The teacher provides conversation prompts or questions to be asked/answered in pairs. When the conversation task is completed, one circle rotates to the next position and new pairs are formed with the same conversational tasks to be completed. The three modeling videos (presented to Groups A and B) demonstrated how to instruct students to execute the inside-outside circles activity and included such conversation topics as discussing the features of a new website and providing suggestions for improving the website’s functioning and look, discussing ways in which some people celebrate Christmas, and discussing what students would like to do if they visited another country (the United Kingdom). During this week, Group C viewed an unrelated video of equivalent length that was a TED talk that focused on recognizing patterns in education as social constructions and identifying areas where these patterns are no longer productive.

**Procedures.** In week 1, prior to viewing their videos and answering a post-video item, all three groups were administered the Teachers Sense of Efficacy Scale (TSES – Short Form) (Appendix B) to assess their baseline teacher self-efficacy; and administered a causal attribution item with a question stem significant to their undergraduate student experience and the Revised Causal Dimension Scale (CDSII) (Appendix C) to assess their causal attribution orientation. During all six weekly sessions, participants in each of the three groups were presented with a micro-appraisal item only after observing their assigned videos. In week 6, after viewing their last set of videos and after answering their micro-appraisal item, all groups were again administered the TSES – Short Form to assess their post condition teacher-self efficacy.
**Instruments.** Four instruments were used in the first pilot study to measure participants’ general teacher self-efficacy (TSES – Short Form), general causal attributions (a causal attribution item and the CDSII), and task specific teacher self-efficacy (a micro-appraisal item).

*TSES – Short Form.* The Teachers Sense of Efficacy Scale (TSES – Short Form; Tschannen-Moran & Wolfolk Hoy, 2001) (See Appendix B) was used to measure participants’ self-efficacy for using effective instructional strategies at the beginning and the end of the pilot study. The TSES – Short Form is a three-factor measure consisting of 12 items which assess a teacher’s self-efficacy for using effective instructional strategies, managing student conduct and classroom behaviors, and promoting student engagement in their learning. Item stems in the TSES - Short Form begin with one of the following phrases - “How much can you do to …”, “How well can you…” , or “To what extent can you…” and end with phrases referencing general teacher behaviors specific to one of the three factors in the measure. Participants assessed the magnitude of their self-efficacy using a 9-point Likert scale response from 1 to 9 with the following anchors (not at all, very little, some degree, quite a bit, and a great deal). The TSES - Short Form has been claimed to be “superior to previous measures of teacher efficacy in that it has a unified and stable factor structure”, well-aligned with self-efficacy theory, and useful for assessing a wide range of abilities that teachers believe are important to their capacity to be effective teachers (Woolfolk Hoy & Burke Spero, 2005). The initial development of the TSES - Short Form demonstrated high reliability for Student Engagement ($\alpha = .81$), Classroom Management ($\alpha = .86$), Instructional Strategies ($\alpha = .86$, and the Composite TSES – Short Form ($\alpha = .90$) (Tschannen-Moran & Woolfolk Hoy, 2001). The TSES – Short Form has also been demonstrated to be a valid measure of teacher self-efficacy across culturally diverse settings (Klassen et al., 2009).
**Causal attribution item.** A causal attribution question significant to undergraduate student experience was asked to assess participants’ generalized causal attribution orientation. This measure was phrased as follows: “Think about the last time you took an important exam. Please identify the primary reason for your doing well or poorly on the exam.” Item choices were as follows: “ability, difficulty of test, effort, luck, other”. Participants were able to type in any explanation if they chose “other”.

**CDSII.** The Revised Causal Dimension Scale (CDSII; McAuley, Duncan, & Russell, 1992) (Appendix C) is a 9-point Likert measure containing twelve items which assesses an individual’s causal attributions across four dimensions: locus of causality, stability, personal control, and external control. These dimensions are underlying factors of Weiner’s (1986) four causal attributions that individuals may attribute to their successes and failures: their ability, task difficulty, their effort, and luck. The CDSII is a reliable measure with Cronbach’s alphas ranging from .60 to .92 in the four studies conducted in this instrument’s development. Participants in the first pilot study were asked to assess the cause of the scenario presented above in their causal attribution item – “Think about the last time you took an important exam. Please identify the primary reason for your doing well or poorly on the exam”. Participants rated their assessments along a 12 continuums that were anchored by opposite causal statements. For example, Item 3 was phrased as follows: “Is the cause something…permanent ( 9 8 7 6 5 4 3 2 1 ) temporary”.

**Micro-appraisal item.** The micro-appraisal item was always presented as “How confident are you in your current ability to perform the teaching behaviors you just observed?” for all three groups across all of their six sessions in this pilot study. Micro-appraisal items did not refer to specific pedagogical tasks, nor were they administered to participants prior to their video
modeling observation, in order to avoid providing additional treatment information to participants beyond the information they received in their assigned groups.

**Data analysis procedures.** Descriptive statistics were calculated to determine weekly participation rates and accessibility of materials. To address research question 1, a One-way MANOVA was conducted to determine if there were statistically significant differences between groups on micro-appraisals of their self-efficacy for performing the six modeled instructional strategies. To address research question 2, descriptive statistics on mean changes in group subscale self-efficacy ratings were reported and a One-way ANCOVA was conducted to determine a statistically significant difference between groups on the change in self-efficacy for instructional strategies controlling for pre-test scores. Pearson product-moment correlation coefficients were computed to assess the relationship between participants’ mean micro-appraisals for the six modeled teaching behaviors and their change in TSES sub-scales. To address research question 3, a One-way MANOVA was conducted to determine statistically significant differences between participant causal attribution orientation and changes in their TSES sub-scales.

**Results**

Participation rates varied throughout the 6-week study (Table 1). There was a large decrease in participation noted for week 2. This week overlapped with a popular charity fundraising event (Thon) at the university and was also the first occurrence of an error in the Qualtrics email distribution of survey links to participants. A redundant system for emailing links was initiated in week 3 to better ensure link deliveries to participants. Over the course of the study:

<table>
<thead>
<tr>
<th>Week #</th>
<th>Group A</th>
<th>Group B</th>
<th>Group C</th>
<th>Total</th>
</tr>
</thead>
</table>

*Participation Rates*
study, the number of participants who spent the necessary time logged on to Qualtrics in order to view the entire length of their assigned videos decreased. Participants mostly logged on for appropriate times during week 1 sessions, but by week 6 very few participants were logged on for sufficient times. Some participants spent less than 30 seconds on their weekly session. Participants who failed to respond to both the pre- and post-TSES surveys were excluded from analysis of the data.

**Research question 1.** How does observation of video modeling affect the magnitude of teachers’ micro-appraisals of their self-efficacy for performing the modeled teaching behaviors? A One-way MANOVA was conducted to determine statistically significant difference between groups on micro-appraisals of their self-efficacy for performing the six modeled instructional strategies in the three study conditions. There was a statistically significant difference in self-efficacy based on group, $F(12, 72) = 3.95$, $p < .001$; Wilk's $\Lambda = 0.363$, partial $\eta^2 = .40$. Because the assumption of homogeneity of variance-covariance was met, the Tukey HSD post hoc procedure was conducted to determine significant pairwise differences in any of the 6 micro-appraisals. There were significant pairwise differences ($p < .05$) between group A (participants who viewed weekly direct instruction videos and 3 modeling videos) and group C (control group participants who only observed a weekly video unrelated to pedagogical techniques) for their self-efficacy micro-appraisals for performing recasting language error correction, teaching...
semantic gradients, and incorporating task-based learning teaching behaviors with students. Group A (M = 6.60, SD = 1.40) had a significantly higher micro appraisal of their self-efficacy for performing recasting behaviors than group C (M = 5.31, SD = 1.45). Group A (M = 7.20, SD = 1.01) also had a significantly higher micro appraisal of their ability use task-based learning techniques with students then group C (M = 5.25, SD = 1.44). However, group A (M = 6.27, SD = 1.34) had a significantly lower micro appraisal of their self-efficacy for teaching semantic gradients than Group C (M = 7.56, SD = 1.63). There was also a significant pairwise difference (p < .05) between group B (participants who viewed 3 weekly modeling videos but no direct instruction video) (M = 6.92, SD = 0.95) and group C (M = 5.31, SD = 1.45) for micro-appraisals of their self-efficacy for performing recasting language error correction.

**Research question 2.** What is the relationship between teachers’ micro-appraisals for specific modeled teaching behaviors and their overall self-efficacy assessment of the larger subscale to which these modeled behaviors represent? Descriptive statistics are reported in table 2 comparing group sample sizes and subscale self-efficacy mean changes. A One-way ANCOVA

<table>
<thead>
<tr>
<th>Group Mean Self-efficacy Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructional Strategies</td>
</tr>
<tr>
<td>--------------------------</td>
</tr>
<tr>
<td><strong>Group A (n=27)</strong> (Video Modeling + Direct Instruction)</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Change</td>
</tr>
<tr>
<td><strong>Group B (n=22)</strong></td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
</tbody>
</table>
(Video Modeling Only)

<table>
<thead>
<tr>
<th>Group C (n=24) (Control)</th>
<th>Change</th>
<th>Pre</th>
<th>Post</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.60</td>
<td>6.84</td>
<td>6.54</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>0.63</td>
<td>6.97</td>
<td>6.91</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>0.73</td>
<td>6.42</td>
<td>6.83</td>
<td>0.41</td>
</tr>
</tbody>
</table>

was conducted to determine a statistically significant difference between groups on the change in self-efficacy for instructional strategies controlling for pre-test scores and to gain power in determining an effect due to group. The ANCOVA determined that there was not a significant effect of group on change in instructional self-efficacy controlling for pre-test scores, F(2,69) = .151, p = .860, partial η² = .004. No significant effects were noted when One-way ANCOVAs were conducted for student engagement F(2,69) = .279, p = .76 and classroom management F(2,69) = .33, p = .72.

Pearson product-moment correlation coefficients were computed to assess the relationship between participants’ mean micro-appraisals for the six modeled teaching behaviors and their change in TSES sub-scales. A positive correlation was observed between participants’ mean micro-appraisal and increase to their self-efficacy for promoting student engagement \( r(71) = 0.234, p = .046 \) and to their self-efficacy for classroom management \( r(71) = 0.238, p = .042 \). However, there was no significant correlation observed between participants’ mean micro-appraisals and increases to their self-efficacy for carrying out effective instructional strategies \( r(71) = 0.080, p = .502 \).

**Research question 3.** Is there a relationship between causal attributions (i.e., effort, ability, luck, and task difficulty) and changes in the magnitude of teachers’ micro-appraisals of
their self-efficacy for specific behaviors they observe in video modeling? An initial One-way MANOVA was conducted to determine statistically significant differences between participant causal attribution orientation and changes in the magnitude of their micro-appraisals. No significant differences were observed.

Given this result and upon reflection that there is no theoretical support for a prediction for an increase in self-efficacy appraisals for divergent pedagogical approaches, research question 3 was revised for the pilot study as follows: Is there a relationship between causal attributions (i.e., effort, ability, luck, and task difficulty) and changes in teachers’ TSES sub-scale appraisals? To test this revision, a One-way MANOVA was conducted to determine statistically significant differences between participant causal attribution orientation and changes in their TSES sub-scales. However, there was not a statistically significant difference in self-efficacy sub-scales based on attribution orientation, $F(12, 72) = 0.66$, $p = .683$; Wilk's $\Lambda = 0.944$, partial $\eta^2 = .03$. The statistical power for this study was .26 for detecting an effect at alpha < .05. There were 51 participants who reported a causal attribution orientation of “effort” (Table 3). The low sample size for participants having other causal attribution orientations would explain the lack of statistical power in this analysis.

<table>
<thead>
<tr>
<th></th>
<th>Group A</th>
<th>Group B</th>
<th>Group C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Difficulty</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>Effort</td>
<td>21</td>
<td>15</td>
<td>15</td>
<td>51</td>
</tr>
<tr>
<td>Luck</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
**Changes for the dissertation study.** As a result of this pilot study, greater care was taken to ensure that study participants received their weekly survey links. Survey links specific to a participant’s group assignment were sent directly to participants via their confirmed university email addresses. The researcher monitored communications with study participants in a consistent and timely manner in order to resolve any materials accessibility issues.

Greater care was also given in the dissertation study to scheduling weekly sessions around important university events and holidays. The six weekly sessions were planned so that no major interruption to participant commitment from student life activities would interfere with participation rates in the full study. Additionally, an incentive of being entered into a lottery to receive an Amazon gift card was also offered for those participants who completed all six of their weekly assigned tasks in their Qualtrics surveys.

Due to the observed decrease in participant times engaged with online materials, measures were taken during the dissertation study to encourage greater participant attention to treatment fidelity such as recruiting participants who showed more interest in second language teaching pedagogy development, informing them that their times logged on to Qualtrics are recorded, and explaining to participants the importance of observing the full length of each video presentation in order for them to receive the full effect of treatment and maintain the integrity of the investigation.

The dissertation study procedure was altered from measuring participants’ micro-appraisals of the six targeted teaching behaviors only after their observation of treatment videos to measuring participants’ micro-appraisals before and after their observation of treatment videos due to a concern that randomly assigned groups may not have equivalent baseline micro-appraisals. The initial posttreatment only procedure was selected for the pilot study to minimize
any presentation of elements of the treatment that may be present in any construction of their pretreatment micro-assessment prompts and also to better distinguish group A (receiving direct instruction and modeling videos) from group B (receiving modeling videos only) in their treatment. Pretest micro-assessment prompts in the pilot study were purposely written generally to be apply to any modeled teaching behavior (i.e., “How confident are you in your current ability to perform the teaching behaviors you just observed?”) so as not to cue participants to the specific salient elements of the targeted teaching behavior and thus act as an additional form of instruction. The dissertation study included a more specific description of the targeted teaching behavior and the pretest administration of the micro-appraisal item was considered part of the treatment for the experimental group. Consequently, the pretest and posttest micro-assessment items were more descriptive so that participants could be better oriented to salient elements of these teaching behaviors that may be novel to them before they were exposed to these behaviors modeled in their assigned videos. As a result of these changes, participants in the experimental group of the dissertation study were given a three-step treatment package which consisted of reading a specific definition of the targeted teaching skill in the pretest micro-appraisal, viewing a direct instruction video which defined the skill and explained the individual steps necessary to execute the skill, and viewing three videos containing different models demonstrating the skill. Consequently, the first research question was altered to address changing the treatment from the observation of modeled videos to the three-step treatment package as described above. The revised first research question is as follows: How does a three-step instructional method which includes: 1) reading the definition of novel skills, 2) viewing a direct instruction video about how to perform the skills, and 3) viewing three videos of models demonstrating the skills in authentic
contexts affect the magnitude of teachers’ micro-appraisals of their self-efficacy for performing the modeled teaching behaviors?

Pretest and posttest micro-appraisals were not included in the control group procedure due to their non-relevance to the observation of videos that were unrelated to teaching behaviors. However, both the experimental and control groups were still assessed on global measures of their teaching self-efficacy (TSES) before and after their six weeks of video observations.

Research question 3 was removed from the full study because it is unlikely that any sampling of students at a major university, where the importance of fostering effort attributions in students is emphasized in teacher preparation courses, would produce enough variance in the sample to ensure enough statistical power to adequately detect an effect for attribution assignment. Accordingly, administration of the CDSII was removed from the procedures of the dissertation study.

**Pilot Study 2**

The second pilot study assessed the validity of instrumentation for assessing participant micro-appraisal of their self-efficacy for performing skills they observed modeled in video demonstrations and their assessment of their knowledge of these same skills. Construct validity was assessed using the multitrait-multimethod matrix approach (MTMM; Campbell & Fiske, 1959; Raykov & Marcoulides, 2011).

Evidence for convergent validity was obtained by determining correlations between two observed measures of the same construct. Evidence for convergent validity for the measurement of participant self-efficacy was obtained by determining the correlation between participants’ micro-appraisals of their self-efficacy and their teacher’s appraisals of participants' self-efficacy using the monotrait-heteromethod (MTHM). This convergent validity coefficient is the
correlation between two separate measures (participant and teacher assessments) of the same construct (participant self-efficacy). Evidence for convergent validity for the measurement of participant knowledge was obtained by determining the correlation between participants’ self-appraisals of their knowledge and their teacher’s appraisals of participants' knowledge using the monotrait-heteromethod (MTHM). This convergent validity coefficient is the correlation between two separate measures (participant and teacher assessments) of the same construct (participant knowledge). A finding of high correlation coefficients for either of these two constructs would lend partial support for their construct validity.

Evidence for discriminant validity may be obtained by determining the correlations of two types of coefficients. The first type of discriminant validity is the correlation between the same method of measurement of two different constructs. This type of discriminant validity uses the heterotrait-monomethod (HTMM). Evidence for the validity of student participants to discriminate between the two constructs (their self-efficacy and knowledge) was obtained by determining the correlation between participants’ micro-appraisals of their self-efficacy and their self-appraisals of their knowledge (HTMM). Evidence for the validity of teachers to discriminate between the two constructs (participants’ self-efficacy and knowledge) was obtained by determining the correlation between teachers’ appraisals of participants' self-efficacy and teachers’ appraisals of participants’ knowledge (HTMM). A finding that the HTMM coefficients are lower than the MTHM coefficients would lend partial support for the construct validity of the measures.

The second type of discriminant validity is the correlation between the different methods of measurement of two different constructs. This type of discriminant validity uses the heterotrait-heteromethod (HTHM). Evidence for discriminant validity, using this method, was
obtained by determining the correlation between student participants' micro-appraisals of their self-efficacy and teachers' appraisals of participants' knowledge. Evidence for discriminant validity was also obtained by determining the correlation between student participants' self-appraisals of their knowledge and teachers' appraisals of participants' self-efficacy. A finding that the HTHM coefficients are lower than HTMM and MTHM coefficients would lend additional support for the construct validity of the measures.

If all conditions of the MTMM approach are met as described above, there would be strong MTMM evidence for the construct validity of each of the four assessment measures.

**Method.** Twenty-three participants were selected from a convenience sample of student-teachers and novice teachers in the San Francisco bay area of California. Participants were informed of their rights as human participants in a research study consistent with the rights explained to participants of the first pilot study and under the approval of Penn State's Intuitional Review Board (IRB). Participants were informed of their benefits for participation in this study (i.e., possibly learning new second language teaching methods, increased awareness of their self-efficacy and knowledge, and information concerning the results of this study). Participants could also receive extra credit upon the discretion of their teacher.

Since the second pilot study was a validation study of instrumentation used for the assessment of participants' self-efficacy and knowledge, no pre or post-administration of the TSES was given. Similarly, participants were not assigned to separate treatment groups. All participants observed the same 6-set series of teacher skill modeling videos preceded by a direct instruction video explaining the skill in each set (as with Group A of the first pilot study), see Appendix A for video links. After observing each set of videos, participants were asked to appraise their self-efficacy for their capacity to carry out the skill and to assess their knowledge
level of the skill. Micro-assessment prompts to assess participant self-efficacy were restructured from the first pilot to increase descriptive specificity of the targeted teaching behaviors being assessed. For example, the first prompt to student participants was – “How confident are you in your current ability to use the recasting method of language error correction? (i.e., revise the spoken error in your conversational response to an ESL student)” Participants again appraised the magnitude of their self-efficacy using a 9-point Likert scale response from 1 to 9 with the following anchors (not at all, very little, some degree, quite a bit, and a great deal). Similarly, participants were asked to appraise their knowledge of the targeted teaching behaviors. For example, the first knowledge assessment prompt was – “How knowledgeable are you about the recasting method of language error correction?” The same 9-point Likert scale was used for participant assessment of their knowledge magnitude. Teachers were also asked to appraise each participant’s level of self-efficacy and knowledge for each of the six modeled skills. Prompts for teachers to assess their students’ self-efficacy were constructed as “How confident is your student in their current ability to use the recasting method of language error correction? (i.e., revise the spoken error in his/her conversational response to an ESL student)” . Prompts for teachers to assess their students’ knowledge were written as “How knowledgeable is your student about the recasting method of language error correction?” The same 9-point Likert scale was used for teacher assessment of these two constructs (see Appendix D for a complete list of items and their appraisal scale).

Results. Pearson’s r correlations were calculated on data from student participants’ micro appraisals of their self-efficacy (SSE) for performing each of the six targeted teaching skills and their assessments of their knowledge (SK) of these skills, and from their teacher’s assessments of students’ self-efficacy (TSE) and students’ knowledge (TK). One hundred and thirty-eight
observations were analyzed for each of the two constructs (self-efficacy and knowledge) across both of the methods of measurement (student and teacher assessment) for the 23 participants in the six targeted teaching behaviors. Multitrait-multimethod correlations are reported in table 4.

There was a strong positive correlation between students’ assessment of their self-efficacy (SSE) and their teachers’ assessment of student self-efficacy (TSE), $r = 0.502, p < .001$.

Table 4

*Multitrait-Multimethod Matrix Correlations*

<table>
<thead>
<tr>
<th></th>
<th>SSE</th>
<th>SK</th>
<th>TSE</th>
<th>TK</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>Pearson Correlation</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>138</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SK</td>
<td>Pearson Correlation</td>
<td>.435**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>138</td>
<td>138</td>
<td></td>
</tr>
<tr>
<td>TSE</td>
<td>Pearson Correlation</td>
<td>.502**</td>
<td>.212*</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td>.012</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>138</td>
<td>138</td>
<td>138</td>
</tr>
<tr>
<td>TK</td>
<td>Pearson Correlation</td>
<td>.229**</td>
<td>.425**</td>
<td>.361**</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.007</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>138</td>
<td>138</td>
<td>138</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).
There was a moderate positive correlation between students’ assessment of their knowledge (SK) and their teachers’ assessment of student knowledge (TK), $r = 0.425$, $p < .001$. The strength of these monotrait-heteromethod correlations lend partial support for the construct validity of these two assessments.

There was a moderate positive correlation between students’ assessment of their self-efficacy (SSE) and their assessment of their knowledge (SK), $r = 0.435$, $p < .001$. There was also a moderate positive correlation between the teachers’ assessment of student self-efficacy (TSE) and their assessment of student knowledge (TK), $r = 0.361$, $p < .001$. The moderate strength of the first heterotrait-monomethod correlation (SSE-SK) demonstrated some discriminant validity for students’ ability to distinguish between their self-efficacy and knowledge for targeted teaching behaviors. However, this correlation (SSE-SK, $r = 0.435$) is larger than the correlation between students’ assessment of their knowledge and their teachers’ assessment of student knowledge (SK-TK, $r = 0.425$). This result does not lend support for the discriminant validity of these assessments. However, the second heterotrait-monomethod correlation (TSE-TK) demonstrated some support for the discriminant validity of these assessments in that this correlation ($r = 0.361$) was weaker than both measures of convergent validity (SSE-TSE, $r = 0.502$; SK-TK, $r = 0.425$).

There was a weak positive correlation between students’ assessment of their self-efficacy (SSE) and their teachers’ assessment of student knowledge (TK), $r = 0.229$, $p = .007$. There was also a weak positive correlation between students’ assessment of their knowledge (SK) and their teachers’ assessment of student self-efficacy (TSE), $r = 0.212$, $p = .012$. The relative weakness of
these heterotrait-heteromethod correlations in comparison to the above monotrait-heteromethod and heterotrait-monomethod correlations lends additional support for the discriminant validity of these measures.

Despite the moderate strengths observed in HTMM correlations and students’ assessment of their knowledge (SK) was more highly correlated with their assessment of their self-efficacy (SSE) than with their teachers’ assessment of student knowledge (TK); the multitrait-multimethod matrix (MTMM) analysis generally provided evidence for the convergent and discriminant validity of these four assessments. Therefore, the general format for participant micro-appraisal of their self-efficacy for targeted teaching behaviors was used in the dissertation study. The construction of the micro appraisal prompts took the following pattern: “How confident are you in your current ability to (insert targeted teaching behavior name)? (i.e., insert brief description of the target teaching behavior to orient participants to the nature of the novel skill being assessed).
Chapter 4 – Methodology

The theoretical framework for the dissertation study was based on social cognitive theory’s conceptions of self-efficacy development and change (Bandura, 1986 & 1997). The effects of vicarious observation of models for self-efficacy development in novice teachers provided a focus of this study based on assertions in the literature concerning the primacy of this source of self-efficacy information for teachers in these contexts (Bandura, 1986; Eccles, Midgley, & Adler, 1984; Tschannen-Moran & Woolfolk Hoy, 2006). Schuman and Relihan (1990) emphasize the necessity of pedagogical modeling in language teacher education because teacher educators must explicitly model important instructional techniques in order for pre-service teachers to develop essential teaching skills required for whole language instruction. Each important behavior of the teaching skill must be modeled in proper sequence. Pedagogical modeling can be enhanced when teacher educators think aloud, or cognitively model, to explain their reasoning for using each teaching behavior and how these behaviors are organized to encompass the complete pedagogical skill (Blomberg et al., 2013; van Es & Sherrin, 2008).

The framework of this study was also grounded in the literature examining the use of video-based technology to support student learning. Observational data from video-based interventions (VBI) are helpful in making distinctions about variations in video presentations that impact the effects of video modeling (Mechling, Ayres, Bryant, & Foster, 2014; Rayner, Denholm, & Sigafous, 2009; Spencer, Mechling, & Ivey, 2015). VBI provides a helpful reference for the use of terminology that characterizes important differences in procedures for study design. Blomberg and her colleagues (2013) made a distinction between presenting observational learning materials under learning conditions where cognitive instructional strategies are used versus learning conditions where situative instructional strategies are used.
They found that the two instructional strategies had differential effects on preservice teachers’ reflection patterns and the way they conceptualized instruction. Instructional aids that direct attention toward salient elements in pedagogical skill modeling in videos may serve to decrease cognitive overload as observers may narrow their focus on specific behaviors that are consistent with the defining elements of the targeted skill (see also, Schworm & Renkl, 2007).

**Theoretical Characteristics of the Dissertation Study**

The design of this study was responsive to critical methodological concerns raised in the literature concerning the appropriate means to measure changes to teachers’ self-efficacy and to the findings from the two pilot studies. Proper methods should target measurement of teacher self-efficacy at an appropriate level of specificity (Bandura, 2006a; Gist & Mitchell, 1992; Pajares, 1996). This study asked participants to assess their self-efficacy for the specific behavior which they had observed models perform on a series of brief videos. Participants were prompted with self-efficacy assessment items for the modeled teaching behavior in the form of “How confident are you in your current ability to perform the teaching behaviors you just observed?” Items specifically identified the targeted teaching behaviors by name to focus participants’ attention on the skill being assessed. Participants indicated the magnitude of their self-efficacy using a 9-point Likert scale response from 1 to 9 with the following anchors: not at all, very little, some degree, quite a bit, and a great deal. The precision of the self-efficacy scale was necessary to capture small changes in participant beliefs and is supported by Bandura (2006a). The anchors were also consistent with those provided by the Teachers’ Sense of Efficacy Scale (TSES, Tschannen-Moran & Wolfolk Hoy, 2001).

This study also measured teacher efficacy development very close in time to the information source event (observation of modeled teaching behaviors). Self-efficacy is typically
measured before and after a period of treatment that consists of a multitude of mini-treatments that function to inform an individual’s larger assessments of their self-efficacy for a class of behaviors (e.g., instructional strategies). By measuring an individual’s micro-appraisals directly after the self-efficacy informing event (observation), this study responds to concerns that distal measures of self-efficacy may be less predictive than proximal measures (Bandura, 1997).

Additionally, this study employed repeated measurement of teachers’ self-efficacy as they were developing. Wheatly (2005) suggested that this method will yield data which may provide insight for how the observation of modeled teaching tasks can affect novice teachers’ cognitive development of their efficacy. Pajares (1996) also emphasized the utility of repeated measure designs when he stated:

The now typical procedure of testing multivariate relationships between self-efficacy, other motivation constructs, and performance attainments in causal models is an improvement over less complex analyses, but providing insights regarding the causal influence of self-beliefs will require that these procedures be conducted on longitudinal or repeated measures designs. (p. 567)

The following research design and analyses of the data were employed to answer this study’s two research questions:

1. How does a three-step instructional method which includes: 1) reading the definition of novel skills, 2) viewing a direct instruction video about how to perform the skills, and 3) viewing three videos of models demonstrating the skills in authentic contexts affect the magnitude of teachers’ micro-appraisals of their self-efficacy for performing the modeled teaching behaviors?
2. What is the relationship between teachers’ micro-appraisals for specific modeled teaching behaviors and their overall self-efficacy assessment of the larger sub-scale to which these modeled behaviors represent? (i.e., instructional strategies self-efficacy)

Methods

Prior to data collection, this study (STUDY00003822) was submitted to the Penn State’s Institutional Review Board and was given the determination of “exempt from IRB initial and ongoing review.” Following this determination, participants were recruited from courses at Penn State focusing on second language teaching methods, assessment, and professional practice for student-teachers to prepare them for working with English language learners in their future classrooms. The principle investigator contacted course instructors via email (Appendix E) and by phone to ask permission to recruit students in person in their classrooms and provide research study information to potential participants. Historically approximately 25 to 30 undergraduate and graduate students are enrolled in each of the 13 sections of these courses (APLING 493, CI 280, EDPSY 408 / SPLED 408, and WL ED 400) at Penn State.

Instructors who agreed to allow recruitment in their courses were unable to offer extra credit points to their students for their participation in the study. Consequently, no equivalent alternative activity was necessary to offer students an extra credit opportunity. Since course instructors were unable to offer extra credit as an incentive for participation, the principle investigator offered a monetary compensation in the form of gift cards to randomly selected participants in order to maximize the participation rate. Sixteen Amazon gift cards were purchased at the following reward levels: one $150 gift card, three $50 gift cards, and twelve $20 gift cards. The Student Research Initiation Grant offered by the Pennsylvania State University’s
College of Education paid for participant incentives. Incentives were awarded in compliance with Penn State policies and in coordination with the EPCSE Financial Assistant.

The principle investigator went to each classroom in turn to solicit students’ voluntary participation in the study. Course instructors and teaching assistants were asked to also inform students, who are absent from the class during the day of recruitment, of the research study opportunity. The investigator explained the purpose and design of the study as well as the activities that students could expect during their participation. The investigator also explained their rights as a participant in a human research study. Participants were asked to print their name and email on the study sign-up form (Appendix F) if they chose to volunteer.

Participants. Study participants were selected from a targeted sample of undergraduate and graduate students attending the second language teaching methods courses mentioned above. These courses were targeted for participant recruitment because students in second language teaching methods courses were thought to be more attentive and responsive to the video modeling plus direct instruction treatment interventions addressing second language teaching pedagogy instruction.

Participants’ ages ranged from 18 to 24. The experimental group (n=36) consisted of 28 individuals who identified as female and 8 individuals who identified as male with a mean age of 21.17. The control group (n=34) consisted of 29 individuals who identified as female and 5 individuals who identified as male with a mean age of 20.47.

Informed consent. The informed consent document was placed at the beginning of each Qualtrics survey throughout the study (Appendix G). Participants were given the opportunity not to consent and could opt out of the survey at any time. If participants gave their voluntary consent, they were directed to survey questions and video links for each study session.
Participants were provided with the name and phone number of the Office of Research Protections to address any concerns regarding their rights as research participants and/or concerns regarding their privacy. The informed consent statement explicitly stated their right to withdraw from the study at any time and that they may choose not to answer any survey item. Informed consent was documented by participants checking “yes” acknowledging that they had read and understood the consent statement, and their desire to freely participate in the study. If participants checked “yes” they were directed toward the session survey and video links. If “no” was checked, the survey was halted and participants were brought to an end screen thanking them for their interest and that no further action on their part is required.

**Study design.** Participants were randomly assigned to two groups (an experimental groups and a control group). Group assignment was determined using the “Random Sequence Generator” on RANDOM.ORG. Both groups participated in six weekly sessions on the Qualtrics survey platform. Both groups were asked to supply their demographic information and respond to a pretest measure of their global teaching self-efficacy (TSES) in the first session. Both groups were also asked to respond to a posttest measure of their global teaching self-efficacy (TSES) in the sixth session. In addition to the common pretest / posttest measure of teacher self-efficacy, the treatment group and control groups were assigned the following unique study conditions:

**Experimental group conditions.** Experimental group participants were given a three-step treatment package which consisted of reading a specific definition of the targeted teaching skill in the pretest micro-appraisal, viewing a direct instruction video which defined the skill and explained the individual steps necessary to execute the skill, and viewing three videos containing different models demonstrating the skill. Changes to participant micro-appraisals were assessed in a follow up micro-appraisal item presented after the three-step treatment package.
The procedure for the experimental group is as follows: In each of their six weekly sessions, participants were given a three-step treatment phase where they were asked to respond to a micro-appraisal item which provided them with a definition of the targeted teaching skill and assessed their baseline self-efficacy for using the skill. They then observed a brief video which provided direct instruction on how to perform the targeted teaching behavior. No modeling of the behavior occurred in the direct instruction video. However, the video introduced the targeted teaching skill by providing a description of the skill and how the skill was to be properly performed. Participants in the experimental group later observed authentic performances of the modeled teaching behavior in a series of three brief videos. During the later assessment phase, participants were asked to again rate their self-efficacy for performing the modeled skill (i.e., micro-appraisal), see appendix H for micro-appraisal items and video observation instructions that encompass the three-step treatment phase and assessment phase.

**Control group conditions.** Control group participants were not provided with any elements of the three-step treatment package. They did not observe the direct instruction video or the series of three videos that modeled the weekly targeted teaching behavior. However, participants were asked to observe a video of a length equivalent to the combined videos of the experimental group. Each of the six weekly videos were unrelated to any teaching behavior, but did address content related to education disconnected from any pedagogical focus. Participants in the control group were not asked to respond to any items to assess their self-efficacy for a specific pedagogical skill. After observing their selected video, they were directed to the end of their survey and informed that no further responses were required of them at this time.

**Materials.** Videos were selected from authentic teaching English as a second language (TESL) pedagogy development videos posted on YouTube. Topics were selected for their
overall representativeness of a broad range of well-defined teaching methods found within common TESL curriculums, the ability to have three or more videos to present to participants, and a minimal production standard to minimize any interference to model observation due to video/audio quality.

**Experimental group video topics.** The six selected weekly topics were presented in the following order: recasting, semantic gradients, finger correction, board races, task-based learning, and inside-outside circles (see Appendix A for links to videos). Study participants in the experimental group received explanations to scaffold their understanding of each modeled language teaching skill, from direct instruction presented in authentic videos (usually from one of the same models who demonstrated the targeted skill). The following explanation of terms are offered here to orient readers of this study:

1. **Recasting** requires the teacher to reformulate all or part of the student’s incorrect utterance within the context of an ongoing conversation (Ding, 2012). For example:
   
   Student: “Last night I goed to the movies with my family.”
   
   Teacher: “Ah, you went to the movies with your family last night. I went to the movies last weekend.”

2. **Semantic gradients** allow students to distinguish between fine differences in meaning of related words by organizing them on a continuum that are typically anchored by antonyms on each end (Reading Rockets, n.d.). For example, teachers could present students with familiar words such as “hot and “cold”, then discuss with students about the fine differences of related words such as “cool”,
“tepid”, “warm”, “chilly”, “scorching”, and “freezing” while organizing each new word along the semantic gradient (word meaning continuum).

3. **Finger correction** is a language teaching technique where the individual sounds of a single word, or individual words of a sentence, are assigned to separate fingers. The purpose of this technique is to isolate individual sounds or words for closer inspection, assist students in correcting spoken errors, and then recombine the units into a coherent and correct utterance (Underhill, 2005).

4. **Board races** involve teachers organizing students into teams to complete a task on the board. Individual team members race to the board to complete a small portion of the task then hand off their marker or piece of chalk to their teammates in a relay race as they compete against another team or teams for points (Astbury, 2015). In second language instruction, this usually involves a language production or correction task.

5. **Task-based learning** requires teachers to propose that students complete a task that “requires students to process language pragmatically in order to achieve an outcome that can be evaluated in terms of whether the correct or appropriate propositional content has been conveyed” (Ellis, 2003, p.16). The targeted task requires students to engage in productive and receptive language skills that resembles authentic language usage. For example, students could be given the task of discussing an area of town where there is a traffic problem. In small groups, they could generate three solutions to this problem and report to the class which solution was the cheapest, most innovative, or most environmentally friendly. Then each group could report which solution they recommended to the
city government and why. It is important that teachers design tasks where students have opportunities to use and extend their existing language knowledge and engage in authentic communications in the targeted language (Willis, 2008).

6. *Inside-outside circles* (Kagan, 1994) allow students to speak and listen to many students as they discuss a common topic. Two circles are formed by having half of a class stand in an inner circle facing outward while the same amount of students form an outer circle facing inward. Each student pairs up with another student from the other circle. Students may summarize what they have read to another student or discuss their thoughts on a common topic. The teacher may stand inside both circles for easier monitoring. After a period of time, teachers can ask one of the circles to rotate left or right for each student to pair up with a new partner.

Experimental group videos were copied and edited to separate direct instruction components, explaining how to carry out the behavior, from actual modeling that demonstrated the behavior. Video modeling with narration (VMN) has been demonstrated to impact learner performance and observational experience (Mechling, Ayres, Bryant, & Foster, 2014; Mechling & Collins, 2012; Smith, Ayres, Mechling, & Smith; 2013). Videos were also edited to connect modeling events that were interrupted by direct instruction or other dialogue. Videos were trimmed to minimize participant time required for observation (i.e., extensive modeling events were trimmed to present only there meaningful components and video introductions/credits were removed).

Additional videos were selected to present to the control group (see Appendix A for links to videos) in place of treatment videos. These videos were selected to avoid any presentation of
direct instruction or modeling of any type of teaching pedagogy. Tangentially related videos, discussing issues in education not dealing with pedagogy, were selected for this purpose. This measure was taken to reduce any threats to validity due to difference is participant history based on their time on task being engaged in similar activities. The length of control group videos were equivalent to the experimental groups. Duration of control group videos fall somewhere between the length of the four videos (1 direct instruction video and 3 modeling videos) that were presented to the experimental group. Control group videos were edited to present their topic in its entirety minus the videos introduction and credits, or videos were trimmed at appropriate content/dialogue pause times so as not to present an awkward break to control group participants.

Once edited, videos were uploaded to the investigator’s YouTube page so that video links could be embedded into Qualtrics surveys and viewed directly by participants without needing to leave their survey platform (see Appendix A for weekly topics, video URLs, and video lengths).

**Video presentation summaries.** During each of the six weeks, the treatment group and the control group were presented with different video packages. The experimental received four videos consisting of a direct instruction video, which explained the targeted pedagogical skill, and three modeling videos where the targeted teaching behavior was demonstrated in authentic contexts. The control group was presented with only one video which discussed education issues but did not address any pedagogical techniques. Control group videos were of equivalent length to experimental group video presentations to create similar time-on-task conditions. The content of the video presentations for both groups are described below and links to each of the videos are provided in Appendix A:

**Week 1.** The focus of this week’s pedagogical skill was on recasting. The direct instruction video (presented to the experimental group) defined recasting as “a form of feedback
you can give students when they are answering questions”. The video provides several examples of language errors produced by English language learners (ELLs) that could be addressed by the teacher by revising the spoken error during the flow of their conversation with their students. The video describes several levels of possible teacher response variations that could scaffold teacher recast interventions based on the nature of the language error produced by the ELL student. The three modeling videos (presented to the experimental group) consisted of teachers working in a variety of contexts where they corrected students’ spoken language errors using the recasting method. One video had captioning so that language errors could also be visual identified by study participants. During this week, the control group viewed an unrelated video of equivalent length that focused on education policy research.

**Week 2.** The focus of this week’s pedagogical skill was semantic gradients. The direct instruction video (presented to the experimental group) defined semantic gradients or gradable opposites by presenting two opposite meaning words, common to students, and putting them on far points along a visual word continuum. The video used “hot” and “cold” as an example and filled in points along the continuum with other known degree variations such as “scorching, warm, cool, and freezing” and a word unknown to students - “lukewarm”. The unknown word’s position along the continuum will indicate its meaning to students. The video also described situations in which gradable opposites could be used and presented variations in visually displaying semantic gradient word clusters. The three modeling videos (presented to the experimental group) demonstrated the semantic gradient technique for teaching students subtle word degree discriminations in a variety of contexts such as with adult ELLs, elementary students, and even two students who were using the technique to construct their own semantic gradient continuum using an outside reference. During this week, the control group viewed an
unrelated video of equivalent length that focused on the unique system of education in the United States.

**Week 3.** The focus of this week’s pedagogical skill was the finger correction technique. The direct instruction video (presented to the experimental group) defined finger correction as consisting of five steps to correct an ELL’s spoken language error: Step 1. A pause by the teacher to indicate the error and potential for possible student consideration, Step 2. Ask the student to repeat what he/she spoke to see if the error was self-corrected, Step 3. The teacher “loads” the words onto their fingers by having the student repeat their sentence slowly and the teacher raises each of his/her fingers word-by-word until the sentence is complete, Step 4. Ask the student to repeat the sentence again with the teacher pointing to and pausing at the finger that is paired with the spoken error, Step 5. The teacher could visually illustrate removing, extending, or adding a word or parts of a word that would form a correct sentence and ask the student to alter their sentence structure. Each of these steps provide increasingly scaffolded instruction and would only be necessary up to the step where the ELL was able to make the necessary correction. The three modeling videos (presented to the experimental group) demonstrated the finger correction technique in a variety of teaching contexts including a foreign classroom, with the teacher also referring to visual aids on a whiteboard, and a point-of-view presentation with the teacher alone simulating an online learning context. During this week, the control group viewed an unrelated video of equivalent length that focused on redefining the national school curriculum to target important life skills necessary for students’ future success.

**Week 4.** The focus of this week’s pedagogical skill was the board race technique. The direct instruction video (presented to the experimental group) defined board races as students practicing language recognition and production by forming two teams and running to the board
to write their responses to the teacher’s grammar/vocabulary questions. Students take turns responding by running to the board to write their answer then running back to their team to handoff their markers to one of teammates. The highest scoring team wins and the teacher reviews the accuracy of answers with the class and may also provide extension activities to further practice the targeted language skills. This activity can be used to practice and reinforce language instruction and a board race activity may be more engaging to students than learning language rules or completing deskwork. The three modeling videos (presented to the experimental group) demonstrated the board race technique in varying adult education and elementary learning contexts and covered topics such as comparative adjective construction, pronoun identification and use, and holiday vocabulary recall and production. During this week, the control group viewed an unrelated video of equivalent length that focused on obtaining a master’s degree in higher education administration at a small university.

Week 5. The focus of this week’s pedagogical skill was using the task-based learning method for English language instruction. The direct instruction video (presented to the experimental group) defined the task-based learning method for learning English by “having students collaborate to solve a problem or complete an assigned task” while using necessary conversation skills, grammar, and vocabulary. The teacher provides the necessary language features (i.e., prior knowledge) to the students that they are to use to complete their assigned tasks. The teacher can elaborate on task requirements in a handout of instructions to increase student precision and provide a structure to adhere to so that students can maintain their focus on using targeted language features. The three modeling videos (presented to the experimental group) demonstrated the task-based learning method for English language instruction with different task-based problem contexts such as creating a utopian society and making a map with
location labels of banks, hospitals, etc.; going on a shopping trip with limited funds and trying to maximize group spending; setting up a task-based learning activity and giving ELLs thorough activity instructions. Videos all showed adult learners in a variety of second language learning contexts. During this week, the control group viewed an unrelated video of equivalent length that focused on using data analytics to provide teachers with information that would be supportive of their administrative tasks and provide them with a comprehensive means to capture and analyze data.

Week 6. The focus of this week’s pedagogical skill was promoting student language production through the use of the inside-outside circles method. The direct instruction video (presented to the experimental group) defined using inside-outside circles as having two groups of students face each other in concentric circles standing or sitting at desks. The teacher provides conversation prompts or questions to be asked/answered in pairs. When the conversation task is completed, one circle rotates to the next position and new pairs are formed with the same conversational tasks to be completed. The three modeling videos (presented to the experimental group) demonstrated how to instruct students to execute the inside-outside circles activity and included such conversation topics as discussing the features of a new website and provide suggestions for improving the website’s functioning and look, discussing ways in which some people celebrate Christmas, and discussing what they would like to do if they visited another country (the United Kingdom). During this week, the control group viewed an unrelated video of equivalent length that was a TED talk that focused on recognizing patterns in education as social constructions and identifying areas where these patterns are no longer productive.

Measures. The Teachers Sense of Efficacy Scale (TSES – Short Form; Tschannen-Moran & Wolfolk Hoy, 2001) (See Appendix B) was used to measure participants’ self-efficacy
for using effective instructional strategies at the beginning and the end of the study. The TSES – Short Form is a three-factor measure consisting of 12 items which assess a teacher’s self-efficacy for using effective instructional strategies, managing student conduct and classroom behaviors, and promoting student engagement in their learning. Item stems in the TSES - Short Form begin with one of the following phrases - “How much can you do to …”, “How well can you…”, or “To what extent can you…” and end with phrases referencing general teacher behaviors specific to one of the three factors in the measure. Participants assessed the magnitude of their self-efficacy using a 9-point Likert scale response from 1 to 9 with the following anchors (not at all, very little, some degree, quite a bit, and a great deal). The TSES - Short Form has been claimed to be “superior to previous measures of teacher efficacy in that it has a unified and stable factor structure”, well-aligned with self-efficacy theory, and useful for assessing a wide range of abilities that teachers believe are important to their capacity to be effective teachers (Woolfolk Hoy & Burke Spero, 2005). The initial development of the TSES - Short Form demonstrated high reliability for Student Engagement (α = .81), Classroom Management (α = .86), Instructional Strategies (α = .86, and the Composite TSES – Short Form (α = .90) (Tschannen-Moran & Woolfolk Hoy, 2001). The TSES – Short Form has also been demonstrated to be a valid measure of teacher self-efficacy across culturally diverse settings (Klassen et al., 2009).

Six micro-appraisal items were used to measure experimental group participants’ self-efficacy for using each of the six targeted teaching skills encompassed by the weekly three-step treatments (see Appendix H). Item stems in each of the micro-appraisal items began with the phrase - "How well can you use…” and ended with a phrase which referenced the specific teaching skill encompassed in that week’s three-step treatment. Micro-appraisal items also included a brief description of the skill as part of the experimental group treatment. Participants
again appraised the magnitude of their self-efficacy using a 9-point Likert scale response from 1 to 9 with the following 5 anchors (not at all, very little, some degree, quite a bit, and a great deal) positioned every other choice with 4 unanchored choices positioned in between the anchored choices. Micro-appraisal items were validated in the second pilot study of this dissertation using the multitrait-multimethod matrix approach (MTMM; Campbell & Fiske, 1959; Raykov & Marcoulides, 2011) to determine if participants were able to distinguish their self-efficacy for using the six targeted teaching skills from their knowledge of these skills. Results provided evidence in support of the construct validity if the micro-appraisal items for assessing skill-specific teacher self-efficacy.

**Procedures.** The study was comprised of six weekly sessions where participants in the experimental group responded to a baseline self-efficacy micro-appraisal to assess their confidence for performing the targeted teaching behavior. For example, the first week’s pretest micro-appraisal item was phrased as “How well can you use the finger correction method to correct a student’s spoken language error? (i.e., use your fingers to target the position of a student’s language error within their spoken sentence)” (see Appendix H for all pretest items).

Participants in the experimental group and the control group observed the video(s) assigned to their specific group. After video observations, participants in the experimental group responded to a posttest micro-appraisal which assessed their degree of confidence for performing the targeted teaching behavior after receiving their three-step treatment package (i.e., the initial self-efficacy micro-appraisal item (with embedded skill definition), observation of a direct instruction video explaining how to perform the skill, and three modeling videos which demonstrated the skill being performed in authentic teaching contexts). For example, the first week’s posttest item was phrased as “After observing the recasting teaching method
demonstrated in the videos: How well can you use the recasting method of language error correction? (i.e., revise the spoken error in your conversational response to an ESL student)” (see Appendix H for all posttest items). The control group did not receive any micro-appraisal items concerning their self-efficacy before or after observing their assigned video unrelated to any teaching methodology.

Baseline data on participants’ global assessment of teaching self-efficacy were collected in week 1 surveys for both groups using the Teachers Sense of Efficacy Scale (TSES – Short Form; Tschannen-Moran & Wolfolk Hoy, 2001) (Appendix B). Participants completed this instrument prior to their first video observations. This instrument was again presented in week six surveys to all groups after their video observations to determine any changes to participants’ global teacher self-efficacy appraisals.

Both groups completed their unique survey items on Qualtrics along with viewing their group specific embedded videos. Participants were not required to follow any links outside of their survey to observe any videos or respond to any question prompts. Twelve unique surveys were constructed in total. The structure of each group’s weekly activities that they were asked to complete are enumerated in table 5 to contrast differences between the conditions of the experimental and control groups, and to illustrate that pre-test data and post-test data are collected outside of the sequence of video observations. Control group participants were not asked make any micro appraisals of their self-efficacy or skill knowledge as this study design purposely omits any pedagogical instruction in control group video presentations.

Table 5

*Weekly Group Activities*

<table>
<thead>
<tr>
<th></th>
<th>Experimental Group</th>
<th>Control Group</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Week 1</th>
<th>Assessment</th>
<th>Treatment (three-steps)</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1. Assessment of their global teaching</td>
<td>2. Step 1: Baseline micro-appraisal of</td>
<td>1. Assessment of their global teaching</td>
</tr>
<tr>
<td></td>
<td>self-efficacy: TSES (Short) pre-test</td>
<td>their self-efficacy for weekly targeted</td>
<td>self-efficacy: TSES (Short) pre-test</td>
</tr>
<tr>
<td></td>
<td></td>
<td>skill with skill definition embedded in item</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Step 2: Observation of a direct</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>instruction video of the targeted teaching</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>behavior</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. Step 3: Observation of three videos</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>modeling the targeted teaching behavior</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assessment</td>
<td>5. Micro self-efficacy appraisal of the</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>modeled teaching behavior</td>
<td></td>
</tr>
<tr>
<td>Weeks</td>
<td>Treatment (three-steps)</td>
<td></td>
<td>Non-treatment Condition</td>
</tr>
<tr>
<td>2, 3, 4,</td>
<td>1. Step 1: Baseline micro-appraisal of</td>
<td></td>
<td>1. Observation of one video unrelated to</td>
</tr>
<tr>
<td>&amp; 5</td>
<td>their self-efficacy for weekly targeted</td>
<td></td>
<td>any type of teaching behavior, but related to</td>
</tr>
<tr>
<td></td>
<td>skill with skill definition embedded in item</td>
<td></td>
<td>education.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Non-treatment Condition

1. Observation of one video unrelated to any type of teaching behavior, but related to education.
<table>
<thead>
<tr>
<th>2. Step 2: Observation of a direct instruction video of the targeted teaching behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Step 3: Observation of three videos modeling the targeted teaching behavior</td>
</tr>
</tbody>
</table>

**Assessment**

| 4. Micro self-efficacy appraisal of the modeled teaching behavior |

<table>
<thead>
<tr>
<th>Week 6</th>
<th>Treatment (three-steps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Step 1: Baseline micro-appraisal of their self-efficacy for weekly targeted skill with skill definition embedded in item</td>
<td></td>
</tr>
<tr>
<td>2. Step 2: Observation of a direct instruction video of targeted the teaching</td>
<td></td>
</tr>
<tr>
<td>3. Step 3: Observation of three videos modeling the targeted teaching behavior</td>
<td></td>
</tr>
</tbody>
</table>

**Assessment**

| 4. Micro self-efficacy appraisal of the modeled teaching behavior |
| 5. Assessment of their global teaching self-efficacy: TSES (Short) post-test |

<table>
<thead>
<tr>
<th>Non-treatment Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Observation of one video unrelated to any type of teaching behavior, but related to education.</td>
</tr>
<tr>
<td>2. Assessment of their global teaching self-efficacy: TSES (Short) post-test</td>
</tr>
</tbody>
</table>
The experimental and control groups were provided with links to their Qualtrics surveys each week on Monday mornings at approximately 10:00 am. Participants were sent their weekly links via direct email using the Penn State UCS system. Participants were allowed to complete their surveys at a time of their choosing and were provided with additional time and reminders to maximize participation.

**Data Analysis Procedures for Dissertation**

The study’s first research question, concerning the effect of video model observation of participant micro-appraisals of their self-efficacy, was addressed by conducting a Wilcoxon signed-rank test to compare the experimental group’s pretreatment and posttreatment rank scores on their micro-appraisals of their self-efficacy for performing each of the six targeted teaching skills. This non-parametric test was selected as an alternative to the paired sample t-test because the data were not normally distributed.

The second research question, concerning the relationship between teachers’ micro-appraisals for specific modeled teaching behaviors and their general appraisal of their self-efficacy for using instructional strategies, was addressed by computing a Spearman’s correlation coefficient. Spearman’s correlation was used because of data from participants’ changes in micro-appraisals were not normally distributed. Spearman’s correlation was used to assess the relationship between the changes in the experimental group’s micro-appraisals for each of the six targeted instructional skills with change to their global assessment of their self-efficacy for employing instructional strategies (i.e., changes in their ratings on the TSES instructional strategies subscale).

Prior to computing the Spearman’s correlation, Cronbach’s alpha was computed on pretest TSES scores to determine the reliability of the administration of this instrument for this
study. A Mann-Whitney U test was conducted to determine if the experimental and control groups had similar initial assessments of their baseline instructional strategy self-efficacy (TSES subscale). The Mann-Whitney U test was selected because of the non-normal distribution of the experimental group’s scores. An independent-samples t-tests was conducted to determine if changes to instructional strategy self-efficacy varied significantly by group. The independent-samples t-tests was used because change scores met all testing assumptions. A post hoc power analysis was also conducted to determine if the independent-samples t-test had sufficient power to detect any observed effect size difference between groups on instructional strategy self-efficacy.
Chapter 5 – Results

The purpose of this study was to examine the effects of video modeling on changes to pre-service teacher self-efficacy. Two research questions were considered:

1. How does a three-step instructional method which includes: 1) reading the definition of novel skills, 2) viewing a direct instruction video about how to perform the skills, and 3) viewing three videos of models demonstrating the skills in authentic contexts affect the magnitude of teachers’ micro-appraisals of their self-efficacy for performing the modeled teaching behaviors?

2. What is the relationship between teachers’ micro-appraisals for specific modeled teaching behaviors and their overall self-efficacy assessment of the larger sub-scale to which these modeled behaviors represent? (i.e., self-efficacy for performing instructional strategies)

Results of the comparative analysis on the effect of treatment on participants’ micro-appraisal scores are reported in this study’s determination of the first research question. Results of the correlation analysis of the relationship between changes in participants’ micro-appraisal scores with changes in their instructional strategies score are reported in this study’s determination of the second research question.

Research Question 1

How does a three-step instructional method which includes: 1) reading the definition of novel skills, 2) viewing a direct instruction video about how to perform the skills, and 3) viewing three videos of models demonstrating the skills in authentic contexts affect the magnitude of teachers’ micro-appraisals of their self-efficacy for performing the modeled teaching behaviors?

Self-efficacy of targeted teaching behaviors. A Wilcoxon signed-rank test was conducted to compare the median of the differences between pretest and posttest micro-appraisal scores. The Wilcoxon signed-rank was selected as a non-parametric alternative to the paired samples t-
test because many of the data violated the assumption of normality with a Shapiro-Wilk
significance value below the preset .05 standard (Table 6).

Table 6

*Test of Normality of Micro-Appraisal Data*

<table>
<thead>
<tr>
<th></th>
<th>Shapiro-Wilk Statistic</th>
<th>Statistic</th>
<th>Df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre_MA1</td>
<td>.220</td>
<td>.910</td>
<td>33</td>
<td>.010</td>
</tr>
<tr>
<td>Post_MA1</td>
<td>.171</td>
<td>.917</td>
<td>33</td>
<td>.015</td>
</tr>
<tr>
<td>Pre_MA2</td>
<td>.177</td>
<td>.925</td>
<td>33</td>
<td>.025</td>
</tr>
<tr>
<td>Post_MA2</td>
<td>.192</td>
<td>.917</td>
<td>33</td>
<td>.015</td>
</tr>
<tr>
<td>Pre_MA3</td>
<td>.172</td>
<td>.943</td>
<td>33</td>
<td>.083</td>
</tr>
<tr>
<td>Post_MA3</td>
<td>.223</td>
<td>.881</td>
<td>33</td>
<td>.002</td>
</tr>
<tr>
<td>Pre_MA4</td>
<td>.214</td>
<td>.907</td>
<td>33</td>
<td>.008</td>
</tr>
<tr>
<td>Post_MA4</td>
<td>.306</td>
<td>.832</td>
<td>33</td>
<td>.000</td>
</tr>
<tr>
<td>Pre_MA5</td>
<td>.148</td>
<td>.956</td>
<td>33</td>
<td>.204</td>
</tr>
<tr>
<td>Post_MA5</td>
<td>.205</td>
<td>.889</td>
<td>33</td>
<td>.003</td>
</tr>
<tr>
<td>Pre_MA6</td>
<td>.200</td>
<td>.915</td>
<td>33</td>
<td>.013</td>
</tr>
<tr>
<td>Post_MA6</td>
<td>.207</td>
<td>.884</td>
<td>33</td>
<td>.002</td>
</tr>
</tbody>
</table>

The four assumption for conducting a Wilcoxon signed-rank test were met for the data: 1) the
two samples must be dependent observations, 2) the paired observations demonstrate
independence, 3) dependent variables are continuous since data was obtained from mean scores
measure, and 4) the level of measurement was ordinal since data was obtained from Likert scale
measures (Statistic Solutions, n.d.).

Test statistics for Wilcoxon signed-rank test are reported in table 7. All test statistics fall
Table 7

Test Statistics

<table>
<thead>
<tr>
<th></th>
<th>Post_MA1 - Pre_MA1</th>
<th>Post_MA2 - Pre_MA2</th>
<th>Post_MA3 - Pre_MA3</th>
<th>Post_MA4 - Pre_MA4</th>
<th>Post_MA5 - Pre_MA5</th>
<th>Post_MA6 - Pre_MA6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.005</td>
<td>.000</td>
</tr>
</tbody>
</table>

Based on negative ranks.

below alpha of .05, so we can reject the null hypothesis for all pretest posttest comparisons that there is no difference between pre- and post-median scores. Results of the rank comparisons between pretest micro-appraisals and posttest micro-appraisals (Table 8) demonstrate that

Table 8

Rank Comparisons

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post_MA1 - Pre_MA1</td>
<td>2</td>
<td>18.00</td>
<td>36.00</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>14.23</td>
<td>370.00</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post_MA2 - Pre_MA2</td>
<td>5</td>
<td>10.50</td>
<td>52.50</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>17.61</td>
<td>475.50</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post_MA3 - Pre_MA3</td>
<td>5</td>
<td>10.90</td>
<td>54.50</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>18.64</td>
<td>540.50</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>35</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
participants consistently scored higher on posttest measures of the six micro-appraisal measures (positive ranks) than on the six pretest measures (negative ranks).

Descriptive statistics are reported in table 9 demonstrating positive gains in participant Table 9

Table 9

Descriptive Statistics for Micro-Appraisals of Targeted Teaching Behaviors

<table>
<thead>
<tr>
<th>Micro-Appraisal</th>
<th>Condition</th>
<th>Mean</th>
<th>N</th>
<th>SD</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Recasting</td>
<td>Pre</td>
<td>5.9167</td>
<td>36</td>
<td>1.27335</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post</td>
<td>7.1111</td>
<td>36</td>
<td>1.42984</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \Delta )</td>
<td>1.19444</td>
<td>36</td>
<td>1.36945</td>
<td>.882</td>
</tr>
<tr>
<td>2 Semantic</td>
<td>Pre</td>
<td>5.8611</td>
<td>36</td>
<td>1.45706</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post</td>
<td>7.2778</td>
<td>36</td>
<td>1.20975</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \Delta )</td>
<td>1.41667</td>
<td>36</td>
<td>1.61024</td>
<td>1.056</td>
</tr>
</tbody>
</table>
self-efficacy for all six micro-appraisal ratings and related and effect sizes. Cohen’s $d$ was calculated using the formula: Cohen’s $d = (M_2 - M_1) / SD_{pooled}$ where $SD_{pooled} = \sqrt{((SD_1^2 + SD_2^2) / 2)$. Cohen’s effect size values ($d = .624$ to $1.056$) suggested a moderate to high practical significance for treatment. These results suggest that the three-step treatment package (skill definition, direct instruction, and skill modeling) had positive practical effects on preservice teachers’ micro-appraisals of their self-efficacy. Specifically, results suggest that when novice teachers read definitions of novel skills, receive direct instruction on how to perform the skills, and observe videos where the teaching skills are modeled, their assessment of their self-efficacy for performing the targeted skills increases when posttreatment measurement occurs immediately after the observation period.
Research Question 2

What is the relationship between teachers’ micro-appraisals for specific modeled teaching behaviors and their overall self-efficacy assessment of the larger sub-scale to which these modeled behaviors represent? (i.e., instructional strategies self-efficacy)

Teachers Sense of Efficacy Scale. Complete data were available from all 70 study participants. Descriptive statistics for the baseline administration of the TSES items are reported in table 10 grouped by the instrument’s three subscales (instructional strategies, student engagement and classroom management). All items were scored on a 9-point Likert scale from 1 (nothing) to 9 (a great deal), see Appendix C for complete TSES – Short instrument.

Table 10

TSES Baseline Descriptive Statistics

<table>
<thead>
<tr>
<th>Items by Subscale</th>
<th>Experimental (n=36)</th>
<th>Control (n=34)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>5. To what extent can you craft good questions for your students?</td>
<td>7.389</td>
<td>1.554</td>
</tr>
<tr>
<td>IS 9. How much can you use a variety of assessment strategies?</td>
<td>7.167</td>
<td>1.765</td>
</tr>
<tr>
<td>10. To what extent can you provide an alternative explanation or example when students are confused?</td>
<td>6.917</td>
<td>1.697</td>
</tr>
<tr>
<td>12. How well can you implement alternative strategies in your classroom?</td>
<td>6.750</td>
<td>1.556</td>
</tr>
</tbody>
</table>
2. How much can you do to motivate students who show low interest in school work?  
SE  
3. How much can you do to get students to believe they can do well in school work?  
SE  
4. How much can you do to help your students value learning?  
5. How much can you do to get children to follow classroom rules?  
CM  
7. How much can you do to calm a student who is disruptive or noisy?  
6. How much can you do to control disruptive behavior in the classroom?  
1. How much can you do to control disruptive behavior in the classroom?  
11. How much can you assist families in helping their children do well in school?  
CM

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Cronbach's Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructional Strategies</td>
<td>6.523</td>
<td>1.813</td>
<td>6.912</td>
</tr>
<tr>
<td>Student Engagement</td>
<td>6.889</td>
<td>1.526</td>
<td>6.765</td>
</tr>
<tr>
<td>Classroom Management</td>
<td>7.250</td>
<td>1.519</td>
<td>7.589</td>
</tr>
<tr>
<td>Composite Scale</td>
<td>7.028</td>
<td>1.647</td>
<td>7.117</td>
</tr>
<tr>
<td>Composite Scale</td>
<td>6.472</td>
<td>1.681</td>
<td>6.677</td>
</tr>
<tr>
<td>Composite Scale</td>
<td>6.778</td>
<td>1.606</td>
<td>6.941</td>
</tr>
<tr>
<td>Composite Scale</td>
<td>6.944</td>
<td>1.638</td>
<td>7.147</td>
</tr>
<tr>
<td>Composite Scale</td>
<td>6.778</td>
<td>1.606</td>
<td>6.941</td>
</tr>
<tr>
<td>Composite Scale</td>
<td>7.194</td>
<td>1.470</td>
<td>6.677</td>
</tr>
</tbody>
</table>

**IS = Instructional Strategies, SE = Student Engagement, CM = Classroom Management**

Cronbach’s alpha was used as a measure of internal consistency for the baseline administration of the TSES. Table 11 presents the means, standard deviations, and Cronbach’s alpha reliability coefficients for each of the three subscales and composite scale. Internal Table 11

*TSES Subscale and Composite Reliability*
<table>
<thead>
<tr>
<th></th>
<th>Number of Items</th>
<th>M</th>
<th>SD</th>
<th>Cronbach’s α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructional Strategies</td>
<td>4</td>
<td>7.086</td>
<td>1.851</td>
<td>.908</td>
</tr>
<tr>
<td>Student Engagement</td>
<td>4</td>
<td>6.971</td>
<td>1.851</td>
<td>.894</td>
</tr>
<tr>
<td>Classroom Management</td>
<td>4</td>
<td>6.889</td>
<td>1.864</td>
<td>.921</td>
</tr>
<tr>
<td>Composite TSES</td>
<td>12</td>
<td>6.982</td>
<td>1.855</td>
<td>.956</td>
</tr>
</tbody>
</table>

reliability on all three subscales and composite scales were high based on accepted cutoff standards of test reliability (DeVellis, 2012; Kline, 2000) and were consistent with Cronbach alpha values reported in the instrument’s initial development (Student Engagement = .81, Classroom Management = .86, Instructional Strategies = .86, Composite TSES = .90; Tschannen-Moran & Woolfolk Hoy, 2001).

**Equivalency of experimental and control groups.** Although study participants were randomly assigned to the experimental and control groups, a Mann-Whitney U test was conducted to compare baseline scores of both groups on the instructional strategies subscale of the TSES. The results of Levene’s test indicated that we could fail to reject the null hypothesis of equivalent variances ($F = 2.174, p = .145$). Therefore, the assumption of homogeneity of variance standard was met for the experimental and control group scores on their baseline measure of instructional strategies self-efficacy. However, although the Shapiro-Wilk test for normality indicated that experimental group scores were normally distributed ($W = .951, p = .114$), the null hypothesis of a normal distribution for the control group was rejected ($W = .850, p < .001$). Consequently, the assumption of normality was unsupported and a nonparametric alternative to the independent-samples t-test was selected to compare the equivalence in means between the experimental and control groups.
The Mann-Whitney U test was selected as a viable alternative to compare group means when parametric assumptions are violated. The data met the four assumptions required by the Mann-Whitney U: 1) an ordinal or continuous dependent variable, 2) two categorical independent groups, 3) independence of observations, and 4) a non-normal distribution. Although descriptive statistics indicated that the experimental group’s scores (mean rank = 33.46) were lower than the control group’s scores (mean rank = 37.66), the Mann-Whitney U-value (U = 538.50) was found not be statistically significant (z = -0.868, p = .385). Therefore, the null hypothesis of equivalent means was failed to be rejected, supporting an equivalence of means. This result, along with random group assignment, supports a claim of equivalent baseline scores for instructional strategy self-efficacy for both groups.

**Comparison of changes to instructional strategy self-efficacy.** Descriptive statistics are presented in table 12 to illustrate the changes to instructional strategy self-efficacy ratings for the experimental and control groups. These composite assessments of their instructional strategy self-efficacy was determined from the mean score of the four related subscale questions on the TSES. This instrument was administered in week 1, before the video observation protocol began, and again in week 6, after both groups had ended their video observations.

Table 12

<table>
<thead>
<tr>
<th>Treatment Condition</th>
<th>PreTest</th>
<th>PostTest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
</tr>
<tr>
<td>Experimental Group</td>
<td>36</td>
<td>7.056</td>
</tr>
<tr>
<td>Control Group</td>
<td>34</td>
<td>7.118</td>
</tr>
</tbody>
</table>
Assumptions for conducting an independent-samples t-test were assessed to determine if changes to instructional strategy self-efficacy varied by group. Assumptions to conduct an independent-samples t-test include: 1) an independent categorical variable of two groups, 2) a continuous dependent variable, 3) homogeneity of variance between groups on the dependent variable, and 4) a normal distribution of the dependent variable. The first assumption was met since observations from the experimental and control groups are independent of each other. The second assumption was met since the dependent variable is the change in instructional strategy scores based on pre- and post-ratings on a continuous 9-point Likert scale. The third assumption for the homogeneity of variance was assessed via Levene’s test of equality of variances for the change in instructional strategy self-efficacy scores for the experimental group and the control group ($F = 2.085, p = .153$). The assumption of homogeneity of variance was supported since the significance value was greater than .05 and the null hypothesis of equal variance was failed to be rejected. The fourth assumption for the normality of the distribution of changes in instructional strategy self-efficacy scores was assessed via the Shapiro-Wilk test for normality. Results indicated that a normal distribution in scores were present for the experimental group ($W = .961, p = .230$) and the control group ($W = .945, p = .084$).

Since these assumptions were met for the data, an independent-samples t-test was conducted to determine if changes to instructional strategy self-efficacy varied by group. The independent-samples t-test indicated that gains in instructional strategy self-efficacy ratings were significantly higher for the experimental group ($M_{gain} = .583$, $SD = 1.372$) than for the control group ($M_{gain} = -.103$, $SD = 1.004$), $t(68) = 2.38, p = .020, d_s = .57$. Cohen’s $d$ was calculated using the formula in figure 2. This formula for effect size is often referred with the $s$ subscript.
Figure 2. Cohen’s d formula (Cohen’s $d_s$) (Cohen, 1988) and considers Bessel’s correction for bias in the estimation of the population variance (Aberson, 2010).

A post-hoc power analysis was conducted for the independent-samples t-test using G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007). This analysis was performed to determine if the independent-samples t-test had sufficient power to detect the observed effect ($d_s = .57$) with an alpha error probability set at 0.05. Results indicate that there was insufficient statistical power ($1 - \beta$) (.65) for the independent-samples t-test to detect the observed effect size (Figure 3). A sensitivity power analysis was then conducted to determine a minimal detectable

$$d = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

Figure 3. Post hoc power analysis to determine observed power.
effect (MDE) given the sample sizes of the experimental group (n=36) and control group (n=34). G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007) was used with power (1 - β) set at 0.80 and α = 0.05, two-tailed. The sensitivity power analysis determined that the independent-samples t-test had sufficient statistical power to detect a slightly higher moderate effect size ($d = .68$)(Figure 4).

**Figure 4.** Sensitivity power analysis to determine a minimal detectable effect (MDE)

**Correlation between changes in micro-appraisals and instructional strategy assessments.** There are five assumptions that must be met with respect to conducting a Pearson’s product-moment correlation: 1) variables must be interval or ratio measurements, 2) variables must be normally distributed, 3) a linear relationship must exist between the two variables, 4) outliers should be minimized, and 5) there must be a homoscedasticity of the data (Warner, 2013). The first assumption was met since the independent variable (change in micro-appraisals)
and the dependent variable (change in self-efficacy for instructional strategies) are both assessed on a 9-point Likert scale. However, the second assumption of a normal distribution among variables was not met. Only the $p$-value for change in instructional strategy self-efficacy was larger than the preset alpha of .05 (Table 13). Consequently, we can only fail to reject the null hypothesis of a normal distribution for the change in score for instructional strategy self-efficacy.

All of the six $p$-values were significant ($< .05$) for change in scores for micro-appraisal, so the null hypotheses of a normal distribution were rejected and a normal distribution for all micro-appraisals could not be assumed. Since, the second assumption for conducting a Pearson’s product-moment correlation was not met; a non-parametric correlation assessment was selected.

### Shapiro-Wilk Test of Normality

<table>
<thead>
<tr>
<th></th>
<th>W</th>
<th>Df</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA 1</td>
<td>.927</td>
<td>33</td>
<td>.030</td>
</tr>
<tr>
<td>MA 2</td>
<td>.925</td>
<td>33</td>
<td>.025</td>
</tr>
<tr>
<td>MA 3</td>
<td>.851</td>
<td>33</td>
<td>.000</td>
</tr>
<tr>
<td>MA 4</td>
<td>.872</td>
<td>33</td>
<td>.001</td>
</tr>
<tr>
<td>MA 5</td>
<td>.918</td>
<td>33</td>
<td>.016</td>
</tr>
<tr>
<td>MA 6</td>
<td>.909</td>
<td>33</td>
<td>.009</td>
</tr>
<tr>
<td>IS</td>
<td>.955</td>
<td>33</td>
<td>.186</td>
</tr>
</tbody>
</table>

Note: MA (1-6) = Change in micro-appraisals (1-6), IS = Change in instructional strategies self-efficacy
A Spearman’s correlation was conducted because the data met the two assumptions required for a valid result (Laerd Statistics, n.d.). First, the two variables were measured on an ordinal, interval, or ratio scale. The data from the independent variables (changes in micro-appraisals) and the dependent variable (changes in instructional strategy self-efficacy) were both from a 9-point Likert based scale. Second, scatter plots revealed a monotonic relationship between scores for changes in instructional strategy self-efficacy and the scores for changes in the six micro-appraisals of targeted teaching behaviors. Spearman’s rho correlations are reported in table 14. There was only one significant correlation between a single micro-appraisal change.

Table 14

**Spearman’s rho Correlations**

<table>
<thead>
<tr>
<th></th>
<th>IS Correlation Coefficient</th>
<th>MA 1 Correlation Coefficient</th>
<th>MA 2 Correlation Coefficient</th>
<th>MA 3 Correlation Coefficient</th>
<th>MA 4 Correlation Coefficient</th>
<th>MA 5 Correlation Coefficient</th>
<th>MA 6 Correlation Coefficient</th>
<th>MA Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IS Correlation Coefficient</td>
<td>MA 1 Correlation Coefficient</td>
<td>MA 2 Correlation Coefficient</td>
<td>MA 3 Correlation Coefficient</td>
<td>MA 4 Correlation Coefficient</td>
<td>MA 5 Correlation Coefficient</td>
<td>MA 6 Correlation Coefficient</td>
<td>MA Total</td>
</tr>
<tr>
<td>IS</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correlation Coefficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA 1</td>
<td>.289</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correlation Coefficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>36</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA 2</td>
<td>.187</td>
<td>.454**</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correlation Coefficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA 3</td>
<td>.424*</td>
<td>.450**</td>
<td>.267</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correlation Coefficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA 4</td>
<td>.314</td>
<td>.269</td>
<td>.161</td>
<td>.472**</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correlation Coefficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>34</td>
<td>35</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>MA 5</td>
<td>.111</td>
<td>.051</td>
<td>.265</td>
<td>.457**</td>
<td>.207</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correlation Coefficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>525</td>
<td>.770</td>
<td>.123</td>
<td>.007</td>
<td>.241</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>36</td>
<td>35</td>
<td>35</td>
<td>34</td>
<td>34</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>--------</td>
<td>----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>MA 6</td>
<td>Correlation Coefficient</td>
<td>.303</td>
<td>.209</td>
<td>.348</td>
<td>.160</td>
<td>.488**</td>
<td>.302</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.072</td>
<td>.222</td>
<td>.038</td>
<td>.358</td>
<td>.003</td>
<td>.078</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>36</td>
</tr>
</tbody>
</table>

Note: IS = Change in instructional strategies self-efficacy, MA (1-6) = Change in micro-appraisals (1-6), MA – Total = Combined changes in all six micro-appraisals.

and changes to the experimental group’s instructional strategies self-efficacy. Participant gains for the self-efficacy micro-appraisal for using the finger correction teaching method (MA -3) was moderately correlated with their larger instructional strategies self-efficacy gain ($r_s = .424, p = .011$)

To provide a more stable measure of change in participant micro-appraisals, a cumulative micro-appraisal change score was formed with unit-weighted z-scores of constituent micro-appraisal items. Construction of a cumulative score is appropriate when an overall summation of item responses would characterize a global response to treatment or an overall response across individual constituent items (Warner, 2013). Cumulative scores are appropriate when individual item correlations suggest they are measuring something in common, items are conceptually consistent with one another, and the combination of items appropriately serves the aim of the research (Anglim, 2009; Furr & Bacharach, 2013). An examination of table 14 demonstrates that micro-appraisal gains are sufficiently correlated with one another to warrant summation. The six items are also conceptually aligned with each other since they were constructed to measure methodologies consistent with second language instruction. Construction of a cumulative score is also in consistent with the aim of this dissertation study (i.e., research question 2 and to make meaningful recommendations to teacher education programs).
Raw micro-appraisal gains were converted to z-scores to provide equal weight for constituent items contributions based on variations in standard deviations. The z-scores were summed to create a cumulative score. Z-score micro-appraisal gains and instructional strategies self-efficacy gains were tested to determine if data met the five assumptions required to perform a Pearson’s product-moment correlation: 1) variables must be interval or ratio measurements, 2) variables must be normally distributed, 3) a linear relationship must exist between the two variables, 4) outliers should be minimized, and 5) there must be a homoscedasticity of the data (Warner, 2013). The initial analysis indicated that the data consisted of four outliers that could not be included in a Pearson’s correlation analysis see figures 5 and 6). As a result, four outliers (experimental group participants 8, 27, 31 & 36)

![Figure 5. Outliers for micro-appraisal gains (z-scores).](image-url)
Figure 6. Outliers for instructional strategies self-efficacy gains.

were omitted from further analysis. The remaining four assumptions for conducting a Pearson correlation were tested on the remaining data (n = 32). The first assumption of variables being on an interval or ratio scale was met because gains in micro-appraisals and instructional strategies self-efficacy were both obtained from Likert scale item responses. The second assumption of a normal distribution was tested using the Shapiro-Wilk test of normality and observation of histograms and QQ-plots. Results of the Shapiro-Wilk tests indicated a normal distribution for $z$-scores of micro-appraisal changes ($W = .968, p = .447$) and for instructional strategies self-efficacy change ($W = .983, p = .890$). We can fail to reject the null hypothesis for a normal distribution of both data sets since their $p$-values were greater than the .05 standard. A visual inspection of the histograms for $z$-scores of micro-appraisal changes (Figure 7) and for instructional strategies self-efficacy change (Figure 8) both suggest a normal distribution of data.
Figure 7. Histogram of z-score micro-appraisal changes.

Figure 8. Histogram of instructional strategies self-efficacy change.
Similarly, a visual inspection of the Q-Q plots for z-scores of micro-appraisal changes (Figure 9) and for instructional strategies self-efficacy change (Figure 10) show linearity for both data sets.

**Figure 9.** Q-Q plot of z-score micro-appraisal changes.

**Figure 10.** Q-Q plot of z-score micro-appraisal change.
This linearity and close proximity of data points to the fit lines also indicate a normal distribution of the data. The third assumption for linearity was tested by conducting a visual inspection of a scatter plot of the z-scores of the micro-appraisal changes plotted against the instructional strategies self-efficacy change (Figure 11). Visual inspection suggests linearity of the data.

Figure 11. Scatter plot of z-scores of micro-appraisal changes plotted against instructional strategies self-efficacy change.

The fourth assumption of minimizing outliers was met upon the initial removal of the four identified outliers when the initial test of assumptions was conducted. The fifth assumption of homoscedasticity was tested by regressing the standardized residuals against the dependent variable (instructional strategies self-efficacy change), see figure 12. Visual inspection of the
Figure 12. Scatter plot testing homoscedasticity.

The scatter plot demonstrates a pattern of dots distributed equally above and below the fitted line as it is viewed from left to right. This pattern indicates that the assumption of homoscedasticity was met for the two variables.

Since all necessary assumptions were met, a Pearson’s product-moment correlation was conducted to determine the relationship between cumulative micro-appraisal change and instructional strategies self-efficacy change. Results indicated a moderate correlation between cumulative micro-appraisal change scores and instructional strategies self-efficacy change scores.
(r = .467, p = .007). These results indicated that although most of the individual changes in micro-appraisals (5) were not predictive of a change in instructional strategies self-efficacy, when considered as a cumulative, their standardized grand mean did demonstrate moderate predictive value.

Summary

The first research question was investigated by conducting a Wilcoxon signed-rank test to compare the median of the differences between participants’ pretest and posttest micro-appraisal scores. Results demonstrate that participants consistently scored higher on posttest measures of the six micro-appraisal items than on the pretest measures. Moderate to high effect sizes were observed for the six treatments. These results suggested that the three-step treatment package (skill definition, direct instruction, and skill modeling) had positive practical effects on preservice teachers’ micro-appraisals of their self-efficacy. The second research question was investigated by conducting a Spearman’s correlation to determine the relationship between the participants’ six micro-appraisal change scores and their instructional strategies self-efficacy change score. Results indicated that only one micro-appraisal change score (MA-3, the finger correction teaching method) was significantly related to instructional strategies self-efficacy change. However, the cumulative score of the six micro-appraisal changes was found to be moderately correlated with instructional strategies self-efficacy change. These results suggested that cumulative scores of conceptually congruent items may be more predictive that individual items.
Chapter 6 – Discussion

The purpose of this research was to determine the effects of a three-step instructional method (skill development, direct instruction, and skill modeling on novice teacher self-efficacy development. The review of the literature determined a gap in our understanding of how novice teachers develop their self-efficacy for the specific behaviors they will need to effectively perform in their classrooms and how these specific micro-appraisals may be related to their larger self-efficacy about their ability to effectively use important instructional strategies. In order to bridge this gap in the literature, this study addressed the following two research questions:

1. How does a three-step instructional method which includes: 1) reading the definition of novel skills, 2) viewing a direct instruction video about how to perform the skills, and 3) viewing three videos of models demonstrating the skills in authentic contexts affect the magnitude of teachers’ micro-appraisals of their self-efficacy for performing the modeled teaching behaviors?
2. What is the relationship between teachers’ micro-appraisals for specific modeled teaching behaviors and their overall self-efficacy assessment of the larger sub-scale to which these modeled behaviors represent? (i.e., instructional strategies self-efficacy)

Discussion of Results

Results of this study are reported and, when possible, interpreted in relation to similar studies in the literature examining the impact of teacher professional development programing on changes to teacher self-efficacy. Although the research design of this study is unique to the study of teacher self-efficacy development, interpretation of the results within the context of the existing literature will provide an orientation for understanding the usefulness of the treatment
intervention for teacher education programs wanting to promote their novice teacher self-efficacy development.

**Research question 1.** The first research question addressed changes to participants’ appraisals of their self-efficacy for specific novel teaching behaviors that they learned about in three-step instruction package which included skill definition embedded in a pretest micro-appraisal item of that skill, a direct instruction video explaining how to carry out the skill, and three brief video presentations of models performing the skill in authentic teaching contexts. The six targeted teaching skills were relevant to their future teaching position requirements. Results of a Wilcoxon signed-rank test demonstrated significant gains in participant micro-appraisals of their self-efficacy for all six targeted teaching behaviors. All effects of treatment were statistically significant with a moderate observed effect size for task-based learning and large effect sizes observed for the five other teaching skills. Improvements to novice teachers’ micro-appraisals of specific targeted teaching skills demonstrated moderate to large treatment effects from the receiving the three-step instructional treatment package.

As discussed in the review of the literature, self-efficacy measurement seldom occurs at the level of task specificity that transpired in this study. Although recommended by researchers (Bong, 2006; Gist & Mitchell, 1992; Pajares, 1997), instruments with items measuring task-specific self-efficacy rarely make their way into the literature base. When these measures are reported, it is usually within the context of instrument development and validation (Rooney & Osipw, 1992). Although no direct comparison between this study and existing studies is possible, some comparison to the literature examining the effects of professional development on teacher self-efficacy may be helpful to place the observed effect sizes of this study in context. Tschannen-Moran and McMaster (2009) found comparable effect size increases of one standard
deviation or more above the mean for treatment groups that included model observation. However, their research design examined the effects of increasing the sources of self-efficacy information input across their four treatment groups. The authors did not examine the effects of model observation isolated from other self-efficacy information sources. A more recent study of the effects of an 8-month professional development program on English language teachers self-efficacy development demonstrated only modest gains to teachers’ instructional strategies self-efficacy \( (d = .421) \). Their professional development treatment consisted of weekly observations, feedback and teacher group discussions surrounding the effective use of the communicative language teaching pedagogy (Ortaçtepe & Akyel, 2015). While the duration of their study was much longer than this study’s 6-week duration, their observed effect size for instructional self-efficacy gain was lower than the observed effect size of this study \( (d_s = .57) \). This comparison suggests that similar or larger effect sizes may be achievable with more modest interventions.

**Research question 2.** The second research question examined the relationship between changes to participants’ self-efficacy micro-appraisals of the six targeted teaching skills and changes to their self-efficacy appraisal on the instructional strategies subscale measure on the TSES. An independent-samples t-test indicated that gains in the instructional strategy self-efficacy for the experimental group (\( M_{\text{gain}} = .583, SD = 1.372 \)) were significantly higher than for the control group (\( M_{\text{gain}} = -.103, SD = 1.004 \), \( t(68) = 2.38, p = .020, d_s = .57 \)). Results of the Spearman’s correlation analysis performed indicated a significant medium strength correlation between participants’ changes to their micro-appraisals for using the finger correction technique with changes to their overall instructional self-efficacy. This result indicated that only one of the changes in micro-appraisals was predictive of change to participant instructional self-efficacy. However, when the changes to all six micro-appraisals were standardized and summed, their
cumulative score also demonstrated a significant medium strength correlation with change in participant overall instructional self-efficacy. This result suggests that the combined effects from the observation of video modeling of specific instructional strategies were positively related to participants’ gains in their self-efficacy for implementing instructional strategies.

Although the correlation between the cumulative change in micro-appraisals were only moderately correlated with changes to instructional self-efficacy, this result compares favorably to other studies in the literature that have longer and more comprehensive treatment phases. Some of these studies reported weak to moderate correlations between teacher professional development and general measures of teacher self-efficacy (Darling-Hammond, Chung, & Frelow, 2002; Ingvarson, Meiers, & Beavis, 2005; Ross, 1994; Ross, 2014). Concerning specific TSES subscales, Ross and Bruce (2007) reported a nonsignificant correlation between their short-term professional development program (one full day followed by three 2-hour after-school sessions) and changes to teacher instructional strategy self-efficacy. While the results of studies with intensive professional development interventions suggest that teacher self-efficacy can be positively affected by manipulating multiple sources of teacher self-efficacy information, they do not consider how specific interventions may effect specific realms of teacher self-efficacy (i.e., how manipulation of specific self-efficacy sources affect self-efficacy subscales). The comparative weaker correlations of the intensive professional development programs suggests that generalized treatments may have weaker effects on global measures of teacher self-efficacy than the effects of specific treatments on more specific measures that are more conceptually congruent.

Implications
The unique methodology and results of this study have implications for theory, research and practice. Implications may at times span across these categories since this study relied heavily on theoretical guidance for its design and research results can be explicitly connected to possible applications for teacher professional development.

**Implications for theory.** The design of this study was responsive to four theoretical concerns raised in the literature but rarely attended to by researchers investigating teacher self-efficacy development. Self-efficacy theory, as reviewed in this study, emphasizes: 1) The importance of model observation as a strong source of information for novice teacher self-efficacy development. 2) Great care should be taken to identify the appropriate level of specificity when investigating aspects of self-efficacy that researchers may want to use for prediction or generalization to other populations. 3) Measurement should occur as soon as possible after any intervention or experimental treatment; and 4) Repeated measurement of self-efficacy should be employed.

Results determined by use of this study design provided support for the validation of these four aspects of self-efficacy theory. Previous research into teacher self-efficacy development did not adequately attend to these four theoretical concerns. Consequently, inferences about relationships between teacher professional development and changes to teacher self-efficacy were not well-supported by the validity arguments of their study designs. The stronger connection between this study’s methodologies and tenets of self-efficacy theory related to change, implies that closer alignment to theory may allow researchers to better observe and measure areas of self-efficacy that were previously inadequately researched in the literature. Moreover, this study design allowed us to peek into the “black box” of treatment interventions and curricular designs, which previously had only measured program level effects, to begin to
identify and measure specific interventions that are correlated with teacher self-efficacy development.

This study demonstrated moderate to large effect sizes for improved teacher self-efficacy at the level of task modeled in the videos observed by novice teachers. This study also demonstrated that one micro-appraisal gain and the cumulative micro-appraisal gain were correlated with instructional self-efficacy gain. The combined effects of the six specific teaching practices were also demonstrated to be moderately correlated with an improvement in novice teachers’ instructional self-efficacy. These results are an important step in furthering our understanding of self-efficacy development. The significant and practical changes to teachers’ micro-appraisals noted in this study and the moderate correlation of their cumulative score suggests that teachers may consider perceptions of their self-efficacy for individual tasks when responding to more generalized items on global and subscale self-efficacy measures. These findings support Tschannen-Moran, Woolfolk Hoy, and Hoy’s (1998) model of teacher self-efficacy change where teachers consider the analysis of the task when assessing their self-efficacy. These findings also support Bandura’s (1997) claim that “self-efficacy should be measured in terms of particularized judgments of capability that may vary across realms of activity, different levels of task demands within a given activity domain, and under different situational circumstances” (p. 6).

Implications for research. The methodology used in this study permitted a closer examination of model observation as a source of novice teacher self-efficacy development. The unique research design was able to observe significant effects for teacher self-efficacy change at a more specific level than had been previously studied in the literature of teacher self-efficacy development. Although micro-appraisal items probed participants’ self-efficacy at a more
specific level than general self-efficacy measures or their subscales, great care was taken to maintain a contextual theme (English language teaching methodologies) that supported an imperative self-efficacy-behavior correspondence (Pajares, 1996). Specific methodologies such as these may allow researchers to examine understudied areas of self-efficacy with methods and instrumentation that are more precise and conceptually aligned with theory. The increased precision and theoretical alignment should allow researchers to support their inferences by making stronger cases for the validity of their research designs.

The statistically significant and practical significant findings of this study suggests that additional insights into areas of self-efficacy research may be garnered if methodologies are more responsive to theoretical concerns and better aligned with the focus of the research questions or hypotheses. These insights may emanate from researchers taking a second look at the way they approach their investigations and the instrumentation and methods they choose to employ. Insights into new approaches may occur when researchers notice a poor fit between existing methods, reported results, inferences made, and theoretical tenets. Theoretical and practical alignment of research methodologies may allow researchers to more precisely measure the highly contextualized factors that teachers consider when appraising their self-efficacy.

**Implications for practice.** Cohen’s effect size values for changes to participants’ micro-appraisals suggest moderate to high practical significance for the observation of specific teaching tasks modeled in videos. Irrespective of their correlation with participants’ assessment of their self-efficacy for using instructional strategies, these effect sizes may demonstrate meaningful change to novice teachers’ task-specific efficacy appraisals. Teacher education and professional development programs may supplement their curricula with multistep learning opportunities that
are likely to enhance their student-teachers self-efficacy for implementing the critical pedagogical skills they are likely to use in their classrooms.

The moderate positive correlation that were demonstrated between one of the micro-appraisal gains and instructional self-efficacy gains, as well as the moderate positive correlation between the cumulative micro-appraisal gain and instructional self-efficacy gains, suggests that teacher education programs could utilize specific multistep learning activities that may improve their student-teachers self-efficacy for using instructional strategies. Similarly, if researcher identified other multistep learning activities that are correlated with teachers’ self-efficacy for student engagement and classroom management); these interventions could also be adapted and employed by teacher education programs to enhance their student-teachers’ self-efficacy in these areas as well. Given the aforementioned benefits of increased teacher self-efficacy to important educational processes and outcomes (Ross, 1998; Tschannen-Moran et al., 1998; Zee & Koomen, 2016), practical changes to teacher education curricula, which promote and sustain its development, may be meaningful to novice teachers’ success in their early careers as well as for improving teacher education outcomes.

**Limitations**

There were several possible limitations of this study, which are related to characteristics of the participants and characteristics of the design. These limitations may have affected the data by introducing systematic error serving to underestimate or overestimate true scores.

**Participant characteristics.** Limitations of this study relating to the characteristics of participants include the size of the study’s sample and participants’ possible response bias.

**Small sample size.** A post hoc power analysis revealed that the sample size of the study provided insufficient power to detect the observed moderate effect size in instructional self-
efficacy gains between the experimental and control groups. Although the difference in mean gains between the two groups were statistically significant, the insufficient power of the comparative analysis did not allow for an inference about the observed practical significance of the treatment effect on teachers’ instructional strategies self-efficacy gains.

**Possible response bias.** The moderate to large effect sizes observed for participants’ micro-appraisals of the six modeled teaching tasks may indicate possible social desirability bias (Krosnick & Presser, 2010). Participants had the opportunity to respond to pre- and posttest micro-appraisals within a short time period. Video presentations ranged from 5 minutes and 39 seconds to 17 minutes and 42 seconds. Shorter presentation times may have existed for participants who did not maintain fidelity of treatment by interrupting the full presentation of their videos. Participants would be likely to remember their initial pretest response for the micro-appraisal and could demonstrate a socially desirable increase due to possibly wanting to please the researcher or having the possible belief that they would be more likely to receive compensation (one of the Amazon gift card rewards). It is likely that social desirability bias may have artificially inflated observed mean gains and effect sizes for participant micro-appraisals of the six modeled teaching skills.

**Study design characteristics.** Limitations of this study relating to methods and procedures include the fidelity of treatment and the conceptual nature of the treatment.

**Fidelity of treatment.** The design of this study included independent observation of the videos for experimental and control groups. This procedure was employed to maximize participation and minimize any disruption to the courses where participants were recruited. Study participants were prompted to observe their assigned videos in full; however, an accurate assessment of treatment fidelity was not possible. Recorded times for participant interaction with
surveys on the Qualtrics platform were recorded and demonstrate great variability. Interaction times ranged from 22 seconds to over 48 hours. Some participants may have been responding to their items as quickly as possible, while others may have left their browsers open for days. While most times were consistent with expected times of engagement based on video lengths, no inferences can be made about actual video observation times and time-on-task rates. Fidelity issues may have affected posttest micro-appraisals scores for the experimental group. Consequently, mean gains and effect sizes may not accurately reflect the effects of treatment.

**Conceptual Nature of the treatment.** The conceptual nature of the treatment presentations may also be a limiting factor of this study. The six modeled teaching behaviors that were presented to participants focused on specific teaching strategies. However, the four items on the instructional strategies subscale of the TSES (table 15) do not all focus solely on implementing instruction. Item 12 is most conceptually related to the modeled teaching strategies.

Table 15

*TSES Instructional Strategies Subscale Items*

<table>
<thead>
<tr>
<th>Item</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>To what extent can you craft good questions for your students?</td>
</tr>
<tr>
<td>9</td>
<td>How much can you use a variety of assessment strategies?</td>
</tr>
<tr>
<td>10</td>
<td>To what extent can you provide an alternative explanation or example when students are confused?</td>
</tr>
<tr>
<td>12</td>
<td>How well can you implement alternative strategies in your classroom?</td>
</tr>
</tbody>
</table>

assessed by micro-appraisal items, but other items on the instructional strategies subscale encompass elements of teacher assessment, planning and adaption.
Additionally, while only one micro-appraisal gain (finger correction method) was statistically related ($p = .011$) to instructional strategies self-efficacy gain, three of the five other micro-appraisal gains approached statistical significance: recasting ($p = .087$), board races ($p = .066$), and inside-outside circles ($p = .072$). The authentic nature and varied production quality may have limited participant attention to salient features of the targeted teaching tasks.

The conceptual incongruence between the six micro-appraisal items and the four instructional strategies items may have suppressed the significance and effects of their individual correlations. However, the moderate correlation observed for the cumulative micro-appraisal gains and instructional strategies self-efficacy gain suggests that conceptual congruence may be captured by consideration of their combined effects.

**Suggestions for Future Research**

Based on the discussion of results, implications and limitations of this study; several suggestions for future research are offered. Suggestions for future investigation incorporate replicating the study with a larger sample size, integrating treatment into course designs, increasing the degree of experimental control, and increasing the duration of the longitudinal design.

**Study replication.** The small sample size ($n = 70$) of this study contributed to the low power in the comparative analysis to detect the observed effect size for differences in instructional self-efficacy gains between the experimental and controls groups. A post hoc sensitivity power analysis determined that a larger sample size was needed to detect the observed moderate effect. Study replication is also of interest to validate the findings and support generalization of the findings to other participants and circumstances. This researcher has an existing opportunity for study replication with a larger sample of university students studying
English language teaching methods in a different context. A larger sample size would permit a comparative analysis to have enough power to detect a similar effect size and replication of findings from a different context would allow for a broader generalization of treatment effects.

**Integration of treatment into course design.** The experimental group in this study received treatment disconnected from their classroom experiences. Although participants were recruited from courses with subject content related to developing teaching methods supporting students who use English as their second language, no attempts were made to align this study’s six weekly treatment sessions of video model observation with course learning activities. As reported in this study’s literature review, Blomberg and her colleagues (2013) argue that integrating video learning into preservice teacher instruction may be more effective for achieving learning goals. Increased alignment between video content and their modeled teaching behaviors with approaches to instruction used by teacher educators may provide a focus for future research to determine if this alignment produces differential treatment effects, and could alter the strength of correlation between micro-appraisal gains and instructional strategies self-efficacy gains.

**Experimental control.** The possible limitations to this study’s design suggest that future research investigating novice teacher self-efficacy development may benefit by a greater experimental control to limit issues related to possible poor treatment fidelity. A larger sample may limit concerns for maximizing participation while more control could be exerted over experimental conditions to increase treatment fidelity. Video observations could occur in a laboratory setting or integrated into novice teachers’ authentic classroom learning experiences. These settings would allow for an increased control over standardizing the duration of treatment presentations and possibly participant attention.
**Longitudinal study.** Longitudinal study is needed to determine the strength (i.e., resiliency) of the improvement in novice teachers’ task-specific self-efficacy micro-appraisals. Posttest data from participants in this study were collected immediately after treatment. It is likely that participant responses were influenced by the possible presence of social desirability bias. Additionally, treatment effects may not be stable and participants may experience delayed effects or experience a dip in their self-efficacy for specific tasks as well as for their general instructional self-efficacy over time. Longitudinal designs that extend beyond the treatment phase and that are more sensitive to bias factors may be productive in lending further support for the validity of the intervention.

Longitudinal designs are also needed to observe preservice teachers in their classrooms after they have completed their teacher education or professional development. It is not known if gains in teacher micro-appraisals are related to their future performance of these specific teaching skills, or if they are also generalizable to other conceptually related teaching methods and behaviors.

**Conclusion**

Results of this study provided new insights into teacher self-efficacy development. Teachers may be considering their capacity to adequately perform several specific tasks when prompted by items on global self-efficacy measures. Conceptual congruence between micro-appraisals and commonly used subscale items allows researchers to detect effects of treatment interventions on these particularized judgements of specific task considerations. Methods that are responsive to concerns raised in the literature about measuring self-efficacy change are likely to provide more evidence when constructing arguments of instrumentation validity. The positive relationship observed between changes in teachers’ micro-appraisals for specific teaching tasks
and changes in their broader instructional strategies self-efficacy supports the use of more theoretically grounded methods to observe and measure teacher self-efficacy development. Although only one individual micro-appraisal gain score was statistically related to instructional strategy self-efficacy gain, use of a cumulative gain score in micro-appraisals may hold the most promise for predicting changes in teachers’ instructional strategies self-efficacy. This study is an initial attempt to identify specific teaching methods that are most conceptually related to the global references to instructional strategies used on the TSES. While the initial findings of this study are promising, more work needs to be done to identify specific micro-appraisals that are statistically correlated with larger assessments of self-efficacy. Improving conceptual congruence between specific micro-appraisals and the broader instructional strategies subscale may improve the magnitude of the correlation between their cumulative score gain and instructional strategy score gain.

The results of this study also hold promise for making meaningful recommendations to teacher education and teacher development programs. It may now be possible to make specific recommendations for curricular change that will lead to improving novice teacher self-efficacy development. Recommendations may be made to teacher education programs to construct multistep instructional methods that incorporate varied learning experiences, as in this study’s three-step treatment package, to improve the self-efficacy of their students for effectively using instructional practices in their future classrooms. Recommendations may also be made to integrate the experiences within content and approaches to instruction used by teacher education programs. The highly contextualized nature of self-efficacy, and interpretation of conditions that change these beliefs, indicate that several three-step instructional presentations may be required so that their combined effects (cumulative score) may be more likely to lead to increases in
teacher self-efficacy as higher levels (broader teacher self-efficacy subscales). Further investigation into this area may identify other modeling events that may be more predictive of gains to teacher instructional strategies self-efficacy. Further investigations integrating modeling events (video or live) into novice teachers’ authentic classroom learning experiences are warranted in specific realms of teacher education so that precise recommendations could be made to improve student-teachers’ self-efficacy development. Recommendations made to teacher education programs based on results from studies in similar contexts are more likely to capture specific course content and processes that would be relevant to pedagogical practices fundamental to these disciplines.
References


Bandura, A. (2004). Social cognitive theory for personal and social change by enabling


Bong, M. (2006). Asking the right question: How confident are you that you could successfully
Preform these tasks? In F. Pajares & T. Urdan (Eds.), *Self-efficacy of adolescents* (pp. 287-306). Greenwich, CT: Information Age.


Center for the Study of Teaching and Policy, Wilson, S. M., Floden, R. E., & Ferrini-Mundy, J.


19.


Ingvarson, L., Meiers, M. & Beavis, A. (2005, January 29). Factors affecting the impact of


O’ Sullivan, M. C. (2002). Action research and the transfer of reflective approaches to


Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale, survey research tells us about teacher effects on student achievement: Insights from the Prospects study of elementary schools. Teachers College Record, 104(8), 1525-1567.


Reading Horizons, 31(2), 105-112.


http://dx.doi.org/10.1037/0022-0663.99.2.285


http://dx.doi.org/10.1016/j.tate.2006.11.005


Zee, M., & Koomen, H. M. Y. (2016). Teacher self-efficacy and its effects on classroom process,
student academic adjustment, and teacher well-being: A synthesis of 40 years of research.

## Appendix A

### Weekly Video Presentations

<table>
<thead>
<tr>
<th>1. Recasting</th>
<th>URL</th>
<th>Trimmed Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td><a href="https://www.youtube.com/watch?v=7NjmGrxhKzY">https://www.youtube.com/watch?v=7NjmGrxhKzY</a></td>
<td>3:06</td>
</tr>
<tr>
<td>Model 1</td>
<td><a href="https://www.youtube.com/watch?v=eOk57nrDDhU">https://www.youtube.com/watch?v=eOk57nrDDhU</a></td>
<td>2:08</td>
</tr>
<tr>
<td>Model 2</td>
<td><a href="https://www.youtube.com/watch?v=tN_bMd_-6O8">https://www.youtube.com/watch?v=tN_bMd_-6O8</a></td>
<td>:16</td>
</tr>
<tr>
<td>Model 3</td>
<td><a href="https://www.youtube.com/watch?v=XAl-CdJjHai">https://www.youtube.com/watch?v=XAl-CdJjHai</a></td>
<td>:09</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>2:33 - 5:39</strong></td>
</tr>
<tr>
<td>Control</td>
<td><a href="https://www.youtube.com/watch?v=kV-7CEr0zo0">https://www.youtube.com/watch?v=kV-7CEr0zo0</a></td>
<td>6:00</td>
</tr>
<tr>
<td></td>
<td><strong>Education Policy Impacts</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. Semantic gradients (Gradable Opposites)</th>
<th>URL</th>
<th>Trimmed Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td><a href="https://www.youtube.com/watch?v=8LfQW2cdi70">https://www.youtube.com/watch?v=8LfQW2cdi70</a></td>
<td>1:42</td>
</tr>
<tr>
<td>Model 1</td>
<td>From Above <a href="https://www.youtube.com/watch?v=xk81oKBbaBk">https://www.youtube.com/watch?v=xk81oKBbaBk</a></td>
<td>2:39</td>
</tr>
<tr>
<td>Model 2</td>
<td><a href="https://www.youtube.com/watch?v=wJ5_uOGWB">https://www.youtube.com/watch?v=wJ5_uOGWB</a></td>
<td>5:58</td>
</tr>
<tr>
<td>Model 3</td>
<td><a href="https://www.youtube.com/watch?v=4NJ7-EYLYzk">https://www.youtube.com/watch?v=4NJ7-EYLYzk</a></td>
<td>:51</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>9:28 - 11:10</strong></td>
</tr>
<tr>
<td>Control</td>
<td><a href="https://www.youtube.com/watch?v=xVJFfbb1uFI">https://www.youtube.com/watch?v=xVJFfbb1uFI</a></td>
<td>9:36</td>
</tr>
<tr>
<td></td>
<td><strong>U.S. System of Education - English Language Notes 2</strong></td>
<td></td>
</tr>
<tr>
<td>3. Finger Correction – 5 Steps</td>
<td>URL</td>
<td>Trimmed Length</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----</td>
<td>----------------</td>
</tr>
<tr>
<td>Definition</td>
<td><a href="https://www.youtube.com/watch?v=ga3ntUdcLD0">https://www.youtube.com/watch?v=ga3ntUdcLD0</a></td>
<td>4:01</td>
</tr>
<tr>
<td>Model 1</td>
<td><a href="https://www.youtube.com/watch?v=dFZvntFym88">https://www.youtube.com/watch?v=dFZvntFym88</a></td>
<td>:31</td>
</tr>
<tr>
<td>Model 2</td>
<td><a href="https://www.youtube.com/watch?v=UwlIuffzF1Q">https://www.youtube.com/watch?v=UwlIuffzF1Q</a></td>
<td>2:05</td>
</tr>
<tr>
<td>Model 3</td>
<td><a href="https://www.youtube.com/watch?v=5OHznFthPtI">https://www.youtube.com/watch?v=5OHznFthPtI</a></td>
<td>:17</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>2:53 - 6:54</td>
</tr>
<tr>
<td>Control</td>
<td><a href="https://www.youtube.com/watch?v=AggAD8DbhEL">https://www.youtube.com/watch?v=AggAD8DbhEL</a></td>
<td>3:43</td>
</tr>
<tr>
<td></td>
<td>What's Education For?</td>
<td></td>
</tr>
<tr>
<td>4. Board Races</td>
<td>URL</td>
<td>Trimmed Length</td>
</tr>
<tr>
<td>Definition</td>
<td><a href="https://www.youtube.com/watch?v=Vf-vS14-TH4">https://www.youtube.com/watch?v=Vf-vS14-TH4</a></td>
<td>2:03</td>
</tr>
<tr>
<td>Model 1</td>
<td>From Above</td>
<td>3:26</td>
</tr>
<tr>
<td></td>
<td><a href="https://www.youtube.com/watch?v=DkVhA6nRze0">https://www.youtube.com/watch?v=DkVhA6nRze0</a></td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td><a href="https://www.youtube.com/watch?v=50AcyHrJ6MA">https://www.youtube.com/watch?v=50AcyHrJ6MA</a></td>
<td>8:19</td>
</tr>
<tr>
<td>Model 3</td>
<td><a href="https://www.youtube.com/watch?v=6VJtbXNfoKk">https://www.youtube.com/watch?v=6VJtbXNfoKk</a></td>
<td>2:36</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>14:21 - 16:24</td>
</tr>
<tr>
<td>Control</td>
<td><a href="https://www.youtube.com/watch?v=yL91NPzp-tI">https://www.youtube.com/watch?v=yL91NPzp-tI</a></td>
<td>14:55</td>
</tr>
<tr>
<td></td>
<td>Higher Education Administration Program Video</td>
<td></td>
</tr>
<tr>
<td>5. Task Based Learning</td>
<td>Definition</td>
<td>URL</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="https://www.youtube.com/watch?v=uZv91VJYm2c">https://www.youtube.com/watch?v=uZv91VJYm2c</a></td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td><a href="https://www.youtube.com/watch?v=nX7-JWm_DzE">https://www.youtube.com/watch?v=nX7-JWm_DzE</a></td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td><a href="https://www.youtube.com/watch?v=NdFQ8qFP7Gs">https://www.youtube.com/watch?v=NdFQ8qFP7Gs</a></td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td><a href="https://www.youtube.com/watch?v=w6tp1Hc_HRA">https://www.youtube.com/watch?v=w6tp1Hc_HRA</a></td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td><a href="https://www.youtube.com/watch?v=wTl-k1R3tMc">https://www.youtube.com/watch?v=wTl-k1R3tMc</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Big Data Analytics in Education &amp; Administration</td>
</tr>
<tr>
<td>6. Inside Outside Circles</td>
<td>Definition</td>
<td>URL</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="https://www.youtube.com/watch?v=rcqrB-_65Dk">https://www.youtube.com/watch?v=rcqrB-_65Dk</a></td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td><a href="https://www.youtube.com/watch?v=T57cYR7hxqA">https://www.youtube.com/watch?v=T57cYR7hxqA</a></td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td><a href="https://www.youtube.com/watch?v=7D1ghieV3Hk">https://www.youtube.com/watch?v=7D1ghieV3Hk</a></td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td><a href="https://www.youtube.com/watch?v=pvZaRyCEZX">https://www.youtube.com/watch?v=pvZaRyCEZX</a></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td><a href="https://www.youtube.com/watch?v=Ci1zPJK8vVU">https://www.youtube.com/watch?v=Ci1zPJK8vVU</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Why Education Needs Entrepreneurs</td>
</tr>
</tbody>
</table>
Appendix B

Teachers Sense of Efficacy Scale (TSES) – Short Form

**Instructions**

Please indicate your opinion about each of the questions below by marking any one of the nine responses in the columns on the right side, ranging from (1) "None at all" to (9) "A Great Deal" as each represents a degree on the continuum. Please respond to each of the questions by considering your current ability to do each of the following:

<table>
<thead>
<tr>
<th>Question</th>
<th>Not at All</th>
<th>Very Little</th>
<th>Some Degree</th>
<th>Quite a Bit</th>
<th>A Great Deal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How much can you do to control disruptive behavior in the classroom?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. How much can you do to motivate students who show low interest in school work?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. How much can you do to get students to believe they can do well in school work?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. How much can you do to help your students value learning?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. To what extent can you craft good questions for your students?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. How much can you do to get children to follow classroom rules?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. How much can you do to calm a student who is disruptive or noisy?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. How well can you establish a classroom management system with each group of students?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. How much can you use a variety of assessment strategies?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. To what extent can you provide an alternative explanation or example when students are confused?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. How much can you assist families in helping their children do well in school?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. How well can you implement alternative strategies in your classroom?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix C

Pilot Study 1: Causal Attribution Item and Revised Causal Dimension Scale (CDSII)

Think about the last time you took an important exam. Please identify the primary reason for your doing well or poorly on the exam.

- Ability
- Difficulty of Test
- Effort
- Luck
- Other

Think about the reason you have indicated above. The items below concern your impression or opinion of this cause of your performance. Mark one number for each of the following questions.

Is the cause something:

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. that reflects an aspect of yourself</td>
<td>reflects an aspect of the situation</td>
</tr>
<tr>
<td>2. manageable by you</td>
<td>not manageable by you</td>
</tr>
<tr>
<td>3. permanent</td>
<td>temporary</td>
</tr>
<tr>
<td>4. you can regulate</td>
<td>you cannot regulate</td>
</tr>
<tr>
<td>5. over which others have control</td>
<td>over which others have no control</td>
</tr>
<tr>
<td>6. inside of you</td>
<td>outside of you</td>
</tr>
<tr>
<td>7. stable over time</td>
<td>variable over time</td>
</tr>
<tr>
<td>8. under the power of other people</td>
<td>not under the power of other people</td>
</tr>
<tr>
<td>9. something about you</td>
<td>something about others</td>
</tr>
<tr>
<td>10. over which you have power</td>
<td>over which you have no power</td>
</tr>
<tr>
<td>11. unchangeable</td>
<td>changeable</td>
</tr>
<tr>
<td>12. other people can regulate</td>
<td>other people cannot regulate</td>
</tr>
</tbody>
</table>
## Appendix D

### Pilot Study 2: Items and Appraisal Scale

#### Pilot Study 2 - Items

<table>
<thead>
<tr>
<th>Skill</th>
<th>Recasting</th>
<th>Semantic Gradients</th>
<th>Finger Correction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SSE</strong></td>
<td>How confident are you in your current ability to use the recasting method of language error correction? (i.e., revise the spoken error in your conversational response to an ESL student)</td>
<td>How confident are you in your current ability to use the semantic gradients method of language teaching? (i.e., guide students to develop an understanding of the subtle differences between words in terms of their magnitude)</td>
<td>How confident are you in your current ability to use the finger correction method of language teaching? (i.e., guide students to develop an understanding of the subtle differences between words in terms of their magnitude)</td>
</tr>
<tr>
<td><strong>SK</strong></td>
<td>How knowledgeable are you about the recasting method of language error correction?</td>
<td>How knowledgeable are you about the semantic gradients method of language teaching?</td>
<td>How knowledgeable are you about the finger correction method of language teaching?</td>
</tr>
<tr>
<td><strong>TSE</strong></td>
<td>How confident is your student in their current ability to use the recasting method of language error correction? (i.e., revise the spoken error in his/her conversational response to an ESL student)</td>
<td>How confident is your student in their current ability to use the semantic gradients method of language teaching? (i.e., guide students to develop an understanding of the subtle differences between words in terms of their magnitude)</td>
<td>How confident is your student in their current ability to use the finger correction method of language teaching? (i.e., guide students to develop an understanding of the subtle differences between words in terms of their magnitude)</td>
</tr>
<tr>
<td><strong>TK</strong></td>
<td>How knowledgeable is your student about the recasting method of language error correction?</td>
<td>How knowledgeable is your student about the semantic gradients method of language teaching?</td>
<td>How knowledgeable is your student about the finger correction method of language teaching?</td>
</tr>
<tr>
<td>SSE</td>
<td>How confident are you in your current ability to use the finger correction method to correct a student's spoken language error? (i.e., use your fingers to target the position of a student's language error within their spoken sentence)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SK</td>
<td>How knowledgeable are you about the finger correction method to correct a student's spoken language error?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSE</td>
<td>How confident is your student in their current ability to use the finger correction method to correct a student's spoken language error? (i.e., use your fingers to target the position of a student's language error within their spoken sentence)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TK</td>
<td>How knowledgeable is your student about the finger correction method to correct a student's spoken language error?</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Skill 4</strong></td>
<td><strong>Board Races</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSE</td>
<td>How confident are you in your current ability to use the board race method for English language teaching? (i.e., organize a competitive game where students race to generate answers in English and then provide appropriate correction when the game is finished)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SK</td>
<td>How knowledgeable are you about the board race method for English language teaching?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSE</td>
<td>How confident is your student in their current ability to use the board race method for English language teaching? (i.e., organize a competitive game where students race to generate answers in English and then provide appropriate correction when the game is finished)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TK</td>
<td>How knowledgeable is your student about the board race method for English language teaching?</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Skill 5</strong></td>
<td><strong>Task-Based Learning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSE</td>
<td>How confident are you in your current ability to use the task-based learning method for English language teaching? (i.e., have students practice English language conversation skills in order to complete an assigned task)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SK</td>
<td>How knowledgeable are you about the task-based learning method for English language teaching?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSE</td>
<td>How confident is your student in their current ability to use the task-based learning method for English language teaching? (i.e., have students practice English language conversation skills in order to complete an assigned task)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TK</td>
<td>How knowledgeable is your student about the task-based learning method for English language teaching?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill 6</td>
<td><strong>Inside-Outside Circles</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSE</td>
<td>How confident are you in your current ability to use the inside-outside circles method for English language teaching? (i.e., promoting student language production by giving students a paired task to perform with a partner and then having them rotate to different partners within a circle formation)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SK</td>
<td>How knowledgeable are you about the inside-outside circles method for English language teaching?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSE</td>
<td>How confident is your student in their current ability to use the inside-outside circles method for English language teaching? (i.e., promoting student language production by giving students a paired task to perform with a partner and then having them rotate to different partners within a circle formation)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TK</td>
<td>How knowledgeable is your student about the inside-outside circles method for English language teaching?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SSE = students ratings of their self-efficacy, SK = students rating of their knowledge, TSE = teacher ratings of students’ self-efficacy, and TK = teacher ratings of students’ knowledge
**Pilot Study 2 – Appraisal Scale**

All items were appraised using a 9-point Likert scale response from 1 to 9 with the following anchors (not at all, very little, some degree, quite a bit, and a great deal).

<table>
<thead>
<tr>
<th>not at all</th>
<th>very little</th>
<th>some degree</th>
<th>quite a bit</th>
<th>a great deal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
Appendix E
Recruitment Email to Course Instructors

Dear Dr. ______,

I am a PhD candidate in the Department of Educational Psychology here at Penn State. I am conducting a pilot study to examine the effects of observing teaching behaviors, modeled in videos, on teachers’ self-efficacy development. I would like to be given the opportunity to discuss my study with your students and recruit their voluntary participation. My research design involves the use of video observations along with brief survey questions. Basically, I would like participants to access a Qualtrics survey once a week for six weeks. Their total commitment time would be for approximately 2 hours (i.e. - 30 minutes for weeks 1 & 6, and 15 minutes for weeks 2-5). All weeks will have brief videos followed by a short survey. Weeks 1 & 6 are a bit longer due to pre and post assessments. Please let me know if I could do this and the most convenient time for me to come to your class to recruit participants. Thank you for your consideration.

Best regards,

David Favre, PhD Candidate

Department of Educational Psychology, Counseling, and Special Education
Appendix F

Research Study: Video Modeling Effects on Teacher Self-Efficacy

Voluntary Participant Sign-Up

EDPSY 010: Dr. ____________

Principle Investigator: David Favre

<table>
<thead>
<tr>
<th>Name</th>
<th>Email (used for course)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix G

Informed Consent

Consent for Exempt Research The Pennsylvania State University

Title of Project: Video Modelling Effects on Teacher Self-Efficacy

Principal Investigator: David E. Favre, PhD Candidate

Telephone Number: 814-865-3296

Advisor: Dr. Stephanie Knight

Advisor Telephone Number: (814) 865-2524

You are being invited to volunteer to participate in a research study. This summary explains information about this research.

- The purpose of this study is to determine the effects from observing teaching behaviors modeled on videos on participants’ appraisals of their teaching confidence.
- During each week of this 6-week study, participants will be provided a link to a Qualtrics survey where they will observe 3 brief videos on a specific language teaching skill and respond to a short survey. During the first week, participants will complete a short teacher efficacy survey (TSES – short form) and then observe 3 brief videos demonstrating the first specific language teaching skill. Participants will respond to an item to assess their confidence to effectively perform the demonstrated skill before and after watching the videos. During the second through fifth weeks, participants will again observe 3 brief videos demonstrating specific language teaching skills and respond to an item to assess their confidence before and after watching video series. During the 6th and final week, participants will again observe 3 brief videos on a specific language teaching skill, respond to an item to assess their confidence before and after watching the videos, and then complete a follow-up teacher efficacy survey (TSES – short form). The principle investigator will prompt participants to access the Qualtrics website at the beginning of each of the 6 weeks of the study and participants may logon to the website at their convenience anytime each week.
- Participants may choose not to answer any survey items or questions, and may stop their participation in the study at any time.
All participant records will be stored in a secure manner and be kept confidential. Once the data has been analyzed, participant identifiers will be removed and deleted.

If you have questions or concerns, you may contact David Favre at 814-865-3296. If you have questions regarding your rights as a research subject or concerns regarding your privacy, you may contact the Office for Research Protections at 814-865-1775.

Your participation is voluntary and you may decide to stop at any time. You do not have to answer any questions that you do not want to answer.

Your participation implies your voluntary consent to participate in the research.

Before making the decision about being in this research you should have

- Discussed this research study with an investigator,
- Read the information in this form, and
- Had the opportunity to ask any questions you may have.

By checking "yes" below you will be indicating that you have received this information, have asked any questions you currently have about the research and those questions have been answered. You will receive a copy of the signed and dated form to keep for future reference.

If you agree to take part in this study, please check "yes" in the box below.

I have read and understood the above consent form and desire of my own free will to participate in the Video Modeling Effects on Teacher Self-Efficacy study.
Appendix H  
Experimental Group: Three-Step Treatment and Assessment Phases

1 Recasting  
[Three-step Treatment Phase]  
(Teaching skill explanation embedded in pretest micro-appraisal)  
How well can you use the recasting method of language error correction? (i.e., revise the spoken error in your conversational response to an ESL student) 

(1 Direct Instruction Video and 3 Teaching Skill Modeling Videos)  
Please watch the following videos and **attend to the demonstrations of the recasting method of language error correction**.

[Assessment Phase]  
After observing the recasting teaching method demonstrated in the videos:  
How well can you use the recasting method of language error correction? (i.e., revise the spoken error in your conversational response to an ESL student)

2 Semantic Gradients  
[Three-step Treatment Phase]  
(Teaching skill explanation embedded in pretest micro-appraisal)  
How well can you use the semantic gradients method of language teaching? (i.e., guide students to develop an understanding of the subtle differences between words in terms of their magnitude) 

(1 Direct Instruction Video and 3 Teaching Skill Modeling Videos)  
Please watch the following videos and **attend to the demonstrations of the semantic gradient method of language teaching**.

[Assessment Phase]  
After observing the semantic gradient teaching method demonstrated in the videos:  
How well can you use the semantic gradients method of language teaching? (i.e., guide students to develop an understanding of the subtle differences between words in terms of their magnitude)
3 Finger Correction

[Three-step Treatment Phase]

(Teaching skill explanation embedded in pretest micro-appraisal)
How well can you use the finger correction method to correct a student's spoken language error? (i.e., use your fingers to target the position of a student's language error within their spoken sentence)

(1 Direct Instruction Video and 3 Teaching Skill Modeling Videos)
Please watch the following videos and attend to the demonstrations of the finger correction method of spoken language correction:

[Assessment Phase]
After observing the finger correction method demonstrated in the videos:
How well can you use the finger correction method to correct a student's spoken language error? (i.e., use your fingers to target the position of a student's language error within their spoken sentence)

4 Board Races

[Three-step Treatment Phase]

(Teaching skill explanation embedded in pretest micro-appraisal)
How well can you use the board race method for English language teaching? (i.e., organize a competitive game where students race to generate answers in English and then provide appropriate correction when the game is finished)

(1 Direct Instruction Video and 3 Teaching Skill Modeling Videos)
Please watch the following videos and attend to the demonstrations of the board race method of English language teaching:

[Assessment Phase]
After observing the board race method demonstrated in the videos:
How well can you use the board race method for English language teaching? (i.e., organize a competitive game where students race to generate answers in English and then provide appropriate correction when the game is finished)

5 Task-Based Learning
[Three-step Treatment Phase]
(Teaching skill explanation embedded in pretest micro-appraisal)
How well can you use the task-based learning method for English language teaching? (i.e., have students practice English language conversation skills in order to complete an assigned task)

(1 Direct Instruction Video and 3 Teaching Skill Modeling Videos)
Please watch the following videos and attend to the demonstrations of the task-based learning method of English language teaching:

[Assessment Phase]
After observing the task-based learning method demonstrated in the videos:
How well can you use the task-based learning method for English language teaching? (i.e., have students practice English language conversation skills in order to complete an assigned task)

6 Inside-Out Circle
[Three-step Treatment Phase]
(Teaching skill explanation embedded in pretest micro-appraisal)
How well can you use the inside-outside circles method for English language teaching? (i.e., promoting student language production by giving students a paired task to perform with a partner and then having them rotate to different partners within a circle formation)

(1 Direct Instruction Video and 3 Teaching Skill Modeling Videos)
Please watch the following videos and attend to the demonstrations of the inside-outside circles method of English language teaching:
[Assessment Phase]
After observing the inside-outside circles learning method demonstrated in the video:
How well can you use the inside-outside circles method for English language teaching? (i.e., promoting student language production by giving students a paired task to perform with a partner and then having them rotate to different partners within a circle formation)
Abbreviated Vita
David E. Favre

Education:
Master of Social Work - August 2006
Florida International University, Miami, FL
Bachelor of Science in Applied Psychology - December 1985
Georgia Institute of Technology, Atlanta, GA

Certifications:
Certificate in Online Teaching - 2015
Pennsylvania State University
Graduate School Teaching Certificate - 2014
Pennsylvania State University
Certificate in English Language Teaching to Adults (CELTA) - 2010
Cambridge University

Publications:

Conference Presentations

Teaching:
Instructor - Pennsylvania State University (2011 - 2014)
- Meeting the Instructional Needs of English Language Learners with Special Needs
- Learning and Instruction
Lecturer - Central Taiwan University of Science and Technology (2006 - 2010)