2016

AN INVESTIGATION OF THE RELATIONSHIP BETWEEN THE TEACHERS' SENSE OF EFFICACY SCALE AND PRESCHOOL CHILDREN'S LITERACY OUTCOMES: MULTILEVEL LONGITUDINAL MODELING AND LONGITUDINAL MEASUREMENT INVARIANCE

Caroline J. Gooden
University of Kentucky, cmjgood@gmail.com
Digital Object Identifier: https://doi.org/10.13023/ETD.2016.478

Click here to let us know how access to this document benefits you.
STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student’s advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student’s thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Caroline J. Gooden, Student

Dr. Michael D. Toland, Major Professor

Dr. Kenneth Tyler, Director of Graduate Studies
AN INVESTIGATION OF THE RELATIONSHIP BETWEEN THE TEACHERS' SENSE OF EFFICACY SCALE AND PRESCHOOL CHILDREN'S LITERACY OUTCOMES: MULTILEVEL LONGITUDINAL MODELING AND LONGITUDINAL MEASUREMENT INVARIANCE

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Education Department of Educational, School, and Counseling Psychology at the University of Kentucky

By

Caroline J. Gooden

Lexington, Kentucky

Co-Directors: Dr. Michael D. Toland, Associate Professor of Educational Psychology
And Dr. Fred Danner, Professor of Educational Psychology

Lexington, Kentucky

2016

Copyright © Caroline J. Gooden 2016
AN INVESTIGATION OF THE RELATIONSHIP BETWEEN THE TEACHERS' SENSE OF EFFICACY SCALE AND PRESCHOOL CHILDREN'S LITERACY OUTCOMES: MULTILEVEL LONGITUDINAL MODELING AND LONGITUDINAL MEASUREMENT INVARIANCE

The current study examined associations between teacher characteristics and child literacy outcomes in a Kentucky preschool sample. The study also examined the psychometric properties of the Teachers’ Sense of Efficacy Scale (TSES; Tschannen-Moran & Woolfolk Hoy, 2001), a frequently used measure of teacher self-efficacy. A widely used preschool assessment instrument, Teaching Strategies GOLD® (GOLD; Heroman, Burts, Berke, & Bickart, 2010), measured child literacy progress. Psychometric examination included confirmatory factor analyses (CFAs) and longitudinal measurement invariance (LMI) of TSES scores. Statistical analyses included longitudinal growth modeling of TSES scores and hierarchical linear modeling (HLM) of TSES and child GOLD literacy scores. CFAs provided evidence that a one-factor model of the 12- and 24-item TSES was reasonable for modeling purposes. LMI analyses indicated that metric invariance was tenable for the 12-item TSES. Growth modeling analyses indicated that growth in TSES scores was minimal across one academic year, using the 12- and 24-item TSES. HLM analyses indicated that teachers with early childhood education certification and higher mean TSES scores were associated with higher child literacy growth in the sample. Study recommendations included further psychometric investigation of the TSES, as well as investigation of associations between other teacher characteristics and child literacy scores.

KEYWORDS: Preschool Teacher Self-Efficacy, Child Literacy, Confirmatory Factor Analysis, Longitudinal Measurement Invariance, Hierarchical Linear Modeling
AN INVESTIGATION OF THE RELATIONSHIP BETWEEN THE TEACHERS' SENSE OF EFFICACY SCALE AND PRESCHOOL CHILDREN'S LITERACY OUTCOMES: MULTILEVEL LONGITUDINAL MODELING AND LONGITUDINAL MEASUREMENT INVARIANCE

By

Caroline J. Gooden

Dr. Michael D. Toland
Co-Director of Dissertation

Dr. Fred Danner
Co-Director of Dissertation

Dr. Kenneth M. Tyler
Director of Graduate Studies

December 16, 2016
Date
This work is dedicated in loving memory of my parents, Barbara and Wallace Jones, who taught me the value of education, and my brother Mark; and in honor of my stepmother Helen McLendon; my siblings Karen, Ellen, Sara, Alison, Justin and their partners; and my daughter Sara Gooden.
Acknowledgements

While this project was my independent work, I am indebted to the faculty and colleagues who helped me complete this study. I was honored to have two chairs for my committee, Drs. Michael Toland and Fred Danner, as well as committee members Drs. Kenneth Tyler, Katherine McCormick, Xin Ma, and Keisha Love who patiently taught me about psychometric and practical content while encouraging my passion for early childhood educational psychology. Drs. Greg Smith and Sycarah Fisher provided expertise at my dissertation defense. I would not have begun and finished this process without the support and encouragement of Dr. Beth Rous and the formatting skills of Ms. Patti Singleton.

My dear friends who walked each step of this journey with me include my partner John Kelly Survant, Steve and Ann Rhodes, Sally Roads, Brent Garrett, Jo Bartlett, Michelle Morris, Martha Pearson, Joy Rodgerson, Jan Reeves, Rebecca Groover, Debi Hardin, Mary Kay Sliney, Dan and Kerri Vantreese, Zijia Li, and more. Thank you for believing in me. Lastly, I am grateful for my faith that carried me each step of the way.
# Table of Contents

Acknowledgements ............................................................................................................ iii

List of Tables ..................................................................................................................... vi

List of Figures ................................................................................................................... vii

Chapter 1: Introduction ........................................................................................................1
  Problem Statement .......................................................................................................... 1
    Study of preschool teacher characteristics ............................................................... 1
    Study of preschool child literacy outcomes ............................................................... 2
    Measurement of TSE and the TSES ........................................................................... 2
  Theoretical and Methodological Framework .............................................................. 3
    Social cognitive theory ............................................................................................. 4
    Psychometric framework .......................................................................................... 5
    Statistical framework .............................................................................................. 7

Chapter 2: Review of Literature ..........................................................................................9
  Demographic Teacher Characteristics and Child Literacy Outcomes ......................... 9
    Teacher educational levels ....................................................................................... 9
    Mixed evidence for teacher education ................................................................... 10
    Comparison of findings for teacher education ....................................................... 13
    Teacher certification, years of experience, professional development, race and ethnicity ................................................................................................................... 14
  Psychological Teacher Characteristics and Child Literacy Outcomes ....................... 14
    Self-reported depression, child- or adult-centered beliefs ....................................... 15
    Emotional supportiveness ....................................................................................... 15
    Mixed evidence for emotional supportiveness ....................................................... 17
    Comparison of findings for emotional supportiveness ............................................ 17
    Teacher self-efficacy ............................................................................................... 18
    Positive evidence for older children ....................................................................... 18
    Positive evidence for preschool children ............................................................... 18
  Measurement of Teacher Self-Efficacy ....................................................................... 19
    Rand items’ approach and related measures ........................................................... 19
    Bandura’s approach and related measures ............................................................... 22
    Gaps in TSE Measurement ...................................................................................... 24
List of Tables

Table 1, Teacher Demographics (Percentages or SD in Parentheses) at Each Wave of Data Collection .................................................................28
Table 2, Child Characteristics (Percentages or SD in Parentheses) at Each Wave of Data Collection ...................................................................30
Table 3, Descriptive Data for Children and Teachers (Percentages or SD in Parentheses): HLM Levels 1, 2, and 3 ........................................................41
Table 4, CFA Model Fit Results for 1-Factor, 3-Factor, and Bifactor Models to the 24-item TSES using MLR Estimation .........................................48
Table 5, CFA Model Fit Results for 1-Factor, 3-Factor, and Bifactor Models to the 12-item TSES using MLR Estimation .......................................50
Table 6, 1-Factor, 3-Factor, and Bifactor CFA Loadings for 12- and 24-Item Fall TSES (n = 265) ........................................................................52
Table 7, Longitudinal Measurement Invariance Results for 12-item TSES scores (n = 467) .................................................................54
Table 8, Growth Model Fit Statistics Using Item Scores for 12-item TSES scores (n = 439) .................................................................55
Table 9, Growth Model Fit Statistics Using Item Scores for 24-item TSES scores (n = 439) .................................................................55
Table 10, Latent or Multiple Indicator Growth Model Fit Statistics Using Factor Scores for 12-item TSES scores (n = 439) .......................57
Table 11, Estimates for HLM Analyses of Child Literacy Growth in Preschool for Child Literacy Outcomes .............................................59
Table 12, HLM Unexplained and Explained Variance .................................................................................................................60
List of Figures

Figure 1, One-Factor, Correlated Three-Factor, and Bifactor Model Conceptual Representation of the TSES .................................................................34
Chapter 1: Introduction

“Teachers’ beliefs in their personal efficacy to motivate and promote learning affect the types of learning environments they create and the level of academic progress their students achieve” (Bandura, 1993, p. 117).

Problem Statement

Study of preschool teacher characteristics. Preschool teacher characteristics, including their beliefs, are an important area of study in Educational Psychology, as they have significant, multi-faceted associations with child outcomes (Burchinal et al., 2008; Clarke-Stewart, Vandell, Burchinal, O’Brien, & McCartney, 2002; Pianta et al., 2005). Studies have examined teacher demographic and psychological characteristics in relation to child outcomes. Teacher demographic characteristics of significance include educational levels, certifications, years of experience, professional development (PD), and race and ethnicity (e.g., Early et al., 2005, 2007; National Institute of Child Health and Human Development [NICHD] Early Child Care Research Network [ECCRN] & Duncan, 2003). Studies also examined teacher psychological characteristics such as self-reported depression, child- or adult-centered beliefs, emotional supportiveness, and teacher self-efficacy (TSE) (e.g., Burchinal et al., 2008; Clarke-Stewart et al., 2002; Howes et al., 2008; Mashburn, Hamre, Downer, & Pianta, 2006; Mashburn et al., 2008; Peisner-Feinberg et al., 2001; Pianta et al., 2005).

One of the psychological variables less frequently studied at the preschool level, and of interest in the current study, is TSE. Bandura (1977a, 1986, 1993, 1997) introduced the concept of self-efficacy as the belief in one’s ability to conduct effective action in response to specific environmental conditions. Gibson and Dembo (1984) further defined TSE as a teacher’s belief in his/her ability to effect desired changes in
children’s behaviors and achievement. Whereas teacher self-esteem refers to a teacher’s feelings of worth, TSE reflects a teacher’s belief is his/her ability to respond effectively to educational challenges (Pajares, 1992).

**Study of preschool child literacy outcomes.** As indicated by a review of existing literature, many studies examined linear associations between preschool teacher characteristics and child outcomes, including social and academic outcomes (Burchinal, Cryer, Clifford, & Howes, 2002; Burchinal et al., 2008; Clarke-Stewart et al., 2002; Early et al., 2007; Howes et al., 2008; Pianta et al., 2005). The current study focused on child literacy outcomes due to the national emphasis on child literacy (No Child Left Behind (NCLB) Act of 2002; Pianta, 2003; Shonkoff & Phillips, 2000) and the number of preschool studies that examined teacher characteristics and child literacy (Burchinal et al., 2008; Early et al., 2007; Howes et al., 2008; Pianta et al., 2005). Child literacy measures examined pre-reading, reading, and letter identification skills (Burchinal et al., 2008; Early et al., 2007; Howes et al., 2008). Investigations of linear associations between teacher characteristics and child literacy outcomes are important to inform preschool teacher education and to improve child progress (Pajares, 1992; Trivette, Dunst, Hamby, & Meter, 2012).

**Measurement of TSE and the TSES.** The review of literature identified a related area of need in the field of educational psychology: the measurement of TSE (Guskey, 1981; Pajares, 1992; Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998) and, in particular, the Teachers’ Sense of Efficacy Scale (TSES; Tschannen-Moran & Woolfolk Hoy, 2001). Although studies showed significant linear associations between TSE and literacy outcomes for older children (Armor et al., 1976; Ashton, Olejnik, Crocker, & McAuliffe, 1982; Berman, McLaughlin, Bass, Pauly, & Zellman, 1977), only one study examined
preschool TSE and child literacy outcomes (Guo, Piasta, Justice, & Kaderavek, 2010). As Guo, Justice, Sawyer, and Tompkins (2011) indicated, “Given the apparent value of preschool teachers’ sense of efficacy, it is surprising that research examining teachers’ sense of efficacy remains limited” (p. 961). A variety of instruments measure TSE, ranging from 2 items developed by Rand researchers (Rand items; Armor et al., 1976) to the 24-item TSES (used in the current study). Further studies of the psychometric properties of the TSES are indicated (Tschanzen-Moran & Woolfolk Hoy, 2001), as few studies to date have examined the intended functionality and purposes of this frequently used measure.

In response to identified areas of need, this study examined linear associations between preschool teacher characteristics and child literacy outcomes, as well as measurement properties of the TSES, in a targeted sample of Kentucky state-funded preschools ($J = 265$ fall, $J = 196$ winter, $J = 298$ spring TSES surveys, $n = 12,304$ children with literacy growth scores). The study’s theoretical framework includes social cognitive theory (Bandura, 1977b, 1986). The study’s methodological framework includes psychometric test theory (i.e., factor analysis; see Crocker & Algina, 1986; Jöreskog, 1971; Long, 1983; Reise, Moore, & Haviland, 2010; Reise, Morizot, & Hays, 2007) as well as statistical approaches including longitudinal growth modeling (McArdle, 1988) and HLM (Raudenbush & Bryk, 1986).

**Theoretical and Methodological Framework**

Social cognitive theory (Bandura, 1977b, 1986), as well as psychometric (Crocker & Algina, 1986; Jöreskog, 1971; Long, 1983; Reise et al., 2007, 2010) and statistical frameworks (i.e., McArdle, 1988; Raudenbush & Bryk, 1986), established a foundation
for examining linear associations between preschool teacher characteristics and child literacy outcomes, and for the measurement of TSE.

**Social cognitive theory.** Developmental theorists (Bandura 1977b, 1986, 1997; Guskey & Passaro, 1994; Heider, 1958; Vygotsky, 1930/1978, 1931/1997; White, 1959) provided a framework for examining linear associations between preschool teacher characteristics and child literacy outcomes. Vygotsky advocated a cultural-historical perspective for learning, in which child development resulted from interactive associations among individual, cultural, and historical factors. He described interactions as dynamic processes that reflected the persons and the settings in which they occurred. Teachers provide knowledge that supports and scaffolds child development. Teacher-child relationships are an interactive process called *obuchenie* in which learning occurs for the teacher and the child (Tudge & Scrimsher, 2003). Social cognitive theorists expanded Vygotsky’s interactive view of the educational process.

Social cognitive theorists (Bandura, 1977a, 1977b, 1993, 1997; Guskey & Passaro, 1994; Heider, 1958; White, 1959) emphasized interactive relationships in the learning process. The recognition of associations between teacher efficacy and child outcomes began with Heider (1958) and White (1959). Heider described the importance of belief in one’s ability to affect change in another person’s behavior (the concept of “I can”). White described a theory of motivation that included the intrinsic need to deal with the environment, the satisfaction of which resulted in a feeling of efficacy. Bandura (1977a, 1977b, 1993, 1997), as well as Guskey and Passaro (1994) expanded theoretical discussions to the specific efficacy of teachers. Bandura’s (1977b) theory of triadic reciprocal causation described interactive relationships among individuals, their behaviors, and their environment. Bandura (1977a) specifically described efficacy in
relation to teachers, as an important characteristic that influenced behaviors and classroom environments. Bandura (1977a) and Guskey and Passaro (1994) described TSE as teachers’ belief in their ability to persist in the accomplishment of teaching tasks, even when challenged by children or educational environments. Whereas Bandura (1993) described changes in teacher behaviors as a result of their beliefs, Guskey (1986) suggested that teacher beliefs changed subsequent to instructional behaviors that improved child outcomes. While theorists differ in the sequence of attitudinal and behavioral change, psychologists (Pajares, 1991; Usher & Pajares, 2008) generally agree with Bandura’s (1997) conceptualization of the sources of self-efficacy.

The sources of self-efficacy beliefs include mastery and vicarious experiences, verbal persuasion, and physiological states (Bandura, 1997; Goddard, Hoy, & Woolfolk, 2000; Pajares, 1991; Usher & Pajares, 2008). Mastery experiences include successful accomplishment of instructional goals, such as teachers’ meeting district curricular goals. Vicarious experiences include information gained by observing other persons, such as when co-teachers learn from their colleagues’ experiences. Verbal persuasion includes the influence of others’ spoken opinions and judgments, including feedback given by mentor teachers. Physiological states that affect self-efficacy include teachers’ anxiety, stress, fatigue, and mood. While social cognitive theory frames the conceptual study of TSE, psychometric theory and statistical approaches provide tools with which to examine its measurement in nested educational structures.

**Psychometric framework.** Researchers have encouraged psychometric examination of instruments whose results inform educational policy decisions (Bisceglia, Perlman, Schaack, & Jenkins, 2009; Pianta, Barnett, Burchinal, & Thornberg, 2009; Pianta et al., 2005; Scarr, Eisenberg, & Deater-Deckard, 1994). Psychometric theory
includes classical test theory (CTT; Crocker & Algina, 1986) as well as a factor analysis (FA; Spearman, 1904) framework. CTT is based on the definition that observed scores have true (i.e., an individual’s actual ability on a latent or unobserved variable) and random error (i.e., unrelated to the trait) components. From a traditional item analysis perspective that is strongly connected to CTT, sample-specific information can be obtained about test items using probability proportion correct levels to indicate item difficulty and Pearson product-moment correlation coefficients (also known as corrected-item total correlations) to examine item discrimination levels (Crocker & Algina, 1986; Fan, 1998; Lord & Novick, 1968). Spearman (1904) introduced the factor analytic framework, which was later expanded (see Crocker & Algina, 1986; Lord & Novick, 1968) to examine correlations between factors and to interpret data patterns (Yong & Pearce, 2003). Factor analysis may be confirmatory (CFA; Long, 1983) which examines known factor structures, or exploratory (EFA; see McDonald, 1999) which investigates the factor structure of a measure. CTT and FA frameworks also include examination of correlations between variables, the internal consistency or reliability (α) of measures (Cronbach, 1951), and longitudinal measurement invariance (LMI; Jöreskog, 1971) of instruments.

Measurement invariance within a CFA framework (as begun by Jöreskog, 1971) examines the equivalency of construct understanding by various groups, whereas LMI addresses the equivalency of construct understanding across multiple time points (Vandenberg & Lance, 2000). As Horn and McArdle (1992) stated, “The general question of invariance of measurement equivalence is a logical prerequisite to the evaluation of substantive hypotheses regarding group differences” (p. 117). Van de Schoot, Lughtig, and Hox (2012) recommended LMI analyses for measurement of latent
constructs such as TSE, to examine whether an instrument’s factor loadings, intercepts, and residual variances are equivalent over time. LMI results may indicate whether changes in scores reflect changes in persons rather than changes in the instrument (van de Schoot et al., 2012). Providing evidence for LMI is an important part of determining instrument measurement quality; the inability to find LMI limits accurate interpretation of scores (Stevens, 2009; Vandenberg & Lance, 2000; van de Schoot et al., 2012). In the current study, a CFA framework (Long, 1983) guided examination of the psychometric characteristics of the TSES. This technique involved model comparisons based on hypothesized model structures and allowed for examination of longitudinal measurement invariance (LMI; Jöreskog, 1971).

**Statistical framework.** Statistical frameworks include approaches that allow for examination of growth in test scores over time, and across multiple related groups such as often occur in educational settings. Two approaches within this perspective include longitudinal growth modeling (McArdle, 1988) and HLM (Raudenbush & Bryk, 1986).

Longitudinal growth modeling may include two perspectives. First, growth modeling includes the use of traditional sum scores or total observed scores at each time point to observe change. Second, latent growth modeling (also known as multiple indicator growth modeling [MIGM; Muthén & Muthén, 1998-2015]) uses items from each time point to reflect the latent or unobserved factor. Both approaches examine repeated measures of the same cases across time, and provide initial status and rate of change of the dependent variable for each case. The only difference is that growth modeling models change in the observed variable, whereas latent growth modeling models change in the latent (factor) variable. Some studies of TSE and child literacy
outcomes used longitudinal growth modeling (Howes et al., 2008; NICHD ECCRN & Duncan, 2003).

The application of HLM (Raudenbush & Bryk, 1986) to educational psychology has evolved from fields including sociology (Goldstein, 1986; Mason, Wong, & Entwisle, 1984), biometrics (Singer, 1998; Singer & Willett, 2003), and statistics (Bryk & Raudenbush, 1992; Lindley & Smith, 1972; Raudenbush & Bryk, 1986). Raudenbush and Bryk (1986) described statistical approaches for analysis of nested data such as repeated observations of children within classrooms (as in the current study). HLM accounts for the dependency that exists within scores for persons over time, when groups are nested, or when individual responses relate to others’ responses (Bryk & Raudenbush, 1992; Goldstein, 1986; Hoffman, 2015; Raudenbush & Bryk, 1986, 2002). Analysis of groups of children and teachers cannot assume independence between variables; in fact, such assumptions miss the complexity of interactions that occur between and among individuals and groups (Raudenbush, 1988). HLM permits the examination of variance resulting from each dependent group (e.g., children and teachers). HLM is appropriate with educational analyses to control for potentially correlated responses among child and teacher level groups (Ma, Ma, & Bradley, 2008; Mashburn et al., 2008; Singer, 1998). Many studies of preschool TSE used HLM, with data collected from multiple sites at one or several points in time.
Chapter 2: Review of Literature

This review of literature presents studies that investigated linear associations between preschool teacher characteristics and child literacy outcomes, as well as studies that examined the measurement of TSE. Preschool teacher characteristics included demographic and psychological variables. Teacher demographic characteristics ranged from educational levels (ranging from high school diplomas to graduate degrees), certifications (ranging from Child Development Associate [CDA] to Interdisciplinary Early Childhood Education [IECE]), years of experience, hours of professional development (PD; pre- and in-service training, workshops, and college courses for teaching staff), to race and ethnicity. Teacher psychological variables studied included self-reported depression, child- or adult-centered beliefs, emotional supportiveness, and TSE. For each teacher characteristic studied, findings ranged from positive to negative linear associations between teacher characteristics and child literacy outcomes. The review suggests reasons for differential findings and gives implications for further study. Lastly, the review summarizes studies that investigated the measurement of TSE, especially in reference to the TSES.

Demographic Teacher Characteristics and Child Literacy Outcomes

Teacher educational levels. Many research initiatives examined teacher educational levels as predictors of child literacy outcomes (Early et al., 2005, 2007; Henry et al., 2004; NICHD ECCRN & Duncan, 2003; NICHD Study of Early Child Care and Youth Development [SECCYD], 2006; Zill et al., 2003). Studies have reported mixed findings for linear associations among preschool teacher educational levels and child literacy. In most studies, teacher educational levels included four categories: high
school or General Education Development (GED) diplomas, Associate’s degrees (AA), Bachelor’s (BA/BS) degrees, and advanced degrees.

**Mixed evidence for teacher education.** Studies reported mixed results for linear associations between teacher educational levels and child literacy outcomes. Early et al. (2007) examined linear associations between teacher educational levels and child literacy outcomes for four-year-olds through secondary analyses of six major preschool studies. The studies included the NICHD SECCYD (2006), Early Head Start Follow-Up (EHS; ACF, 2002), Head Start Family and Child Experiences Survey (FACES; Zill et al., 2003), National Center for Early Development and Learning (NCEDL) Multi-State Pre-Kindergarten and State-Wide Early Education Program (Multi-State and SWEEP; Early et al., 2005) studies, and Georgia Early Care Study (GECS; Henry et al., 2004). The NICHD SECCYD (2006) study investigated the influence of childcare quality on child outcomes for children in a geographically diverse sample selected from hospital birth records in 10 sites \((J = 10)\) during 1991 \((n = 1,364\) children at 1 month of age). The EHS study assessed EHS and control children from 17 sites \((J = 17)\) during their prekindergarten year. The FACES study (Zill et al., 2003) examined the quality of Head Start programs, based on children from 63 randomly selected programs \((J = 63)\) during their first year in Head Start. The NCEDL conducted two national studies to examine state-funded preschool program quality and child outcomes, which represented 79% of all children participating in state-funded U.S. preschools at that time (Early et al., 2005). The NCEDL Multi-State study \((J = 238\) sites) assessed preschool programs in six states in 2001 to 2002; the SWEEP study \((J = 463\) sites) examined preschools in five additional states from 2003-2004. The GECS study (Henry et al., 2004) examined children
attending public and private preschools in Georgia, with randomly selected children from one randomly selected classroom at each site \((J = 128 \text{ sites}; n = 630 \text{ children})\).

The six studies administered many different measures of child literacy. Each of the six studies administered the Woodcock-Johnson Test of Achievement (WJ-III; Woodcock, McGrew, & Mather, 2001) Letter Word Identification subtest to measure reading skills. Three of the studies (GCES, NCEDL Multi-State and SWEEP) used the WJ-III subtest version; three (EHS, FACES, NICHD) studies used the earlier Woodcock-Johnson Revised subtest (WJ-R; Woodcock & Johnson, 1990). In addition, the NCEDL studies included three other literacy measures. First, NCEDL study teachers rated children’s literacy skills with the Academic Rating Scale (ARS) from the Early Childhood Longitudinal Study-Birth Cohort (West, Denton, & Germino-Hausken, 2000). The ARS included nine items with 5-point Likert-type response options ranging from 1 (not yet) to 5 (proficient). Second, NCEDL studies included an unstandardized project-developed letter identification measure in which children identified up to 26 upper and lowercase letters (NCEDL, 2001). Third, NCEDL studies administered the WJ-III (Woodcock et al., 2001) Sound Awareness, Rhyming subtest that measures the phonological component of literacy (Mashburn et al., 2008). Further, the NICHD ECCRN study included other measures of child literacy: the Bayley II Scales of Infant Development (Bayley, 1993) and the WJ-R (Woodcock & Johnson, 1990) Incomplete Words Scales subtest (NICHD ECCRN & Duncan, 2003).

In their analyses across the six studies, Early et al. (2007) used regression analyses for all studies and HLM with studies of children nested within classrooms and programs. They found positive, significant linear associations between teachers with BA/BS degrees and child pre-reading scores for GECS \((F[1, 494] = 11.39, p < .001)\) and
NCEDL Multi-State and SWEEP \(F[1,1,475] = 4.81, p < .05\) studies. They found no significant linear associations between teacher education and child literacy outcomes at EHS, FACES, or NICHD sites. However, other analyses of the NICHD study found mixed results for a linear association between teacher education and child cognitive growth, which included pre-reading scores (NICHD ECCRN & Duncan, 2003).

NICHD ECCRN and Duncan (2003) reported different results than did Early et al. (2007), based on growth modeling of NICHD child data. NICHD ECCRN and Duncan (2003) conducted a 3-level growth model to examine the influence of childcare quality (including teacher educational levels) on child achievement (which included pre-reading skills) scores at two time points. They conducted two types of analyses (including status and growth models) for child literacy scores at 24 \((n = 887)\) and 54 months \(n = 985\) to estimate the influence of early and later program quality. Measures of child literacy included the Bayley II (Bayley, 1993) for children at 24 months of age, and the WJ-R (Woodcock & Johnson, 1990) Letter Word Identification and Incomplete Words Scales subtests for children at 54 months of age. Status model analyses (as recommended by Blau, 1999) assumed that preschool children’s scores at 24 and at 54 months reflected child, family, and childcare characteristics at each wave. Growth model analyses (Cronbach & Furby, 1970; Diggle, Liang, & Zeger, 1994) used the preschool 54-month scores as outcome variables, with changes in child scores from 24 to 54 months included in each model. For each type of analysis, the researchers fits three models as follows: Model 1 included center type and hours per week; Model 2 added child ethnicity and gender and maternal education; and Model 3 added family income, home environment, and child temperament variables. For status model analyses, teacher educational levels were not associated with child literacy outcomes when examining 54-month scores only.
(Model 1 estimate = 0.61, $p < .05$; Model 2 estimate $ns$; Model 3 estimate $ns$). However, teacher educational levels were positively, linearly associated with child literacy outcomes at 54 months when conducting growth model analyses between adjacent times points (i.e., score differences from 24 to 54 months; Model 1 estimate = 0.50, $p < .05$; Model 2 estimate = 0.64, $p < .01$; and Model 3 estimate = 0.65, $p < .01$).

In comparison to Early et al.’s (2007) finding that teacher education was a modest predictor of child literacy outcomes for three of the six (GCES, NCEDL Multi-State and SWEEP) studies, Howes et al. (2008) reported different results using HLM analyses of an unstandardized literacy measure for the two NCEDL studies. Howes et al. examined results for a project-developed letter identification (literacy) measure, in which children identified up to 26 upper and lowercase letters (NCEDL, 2001). They found no linear association between teacher educational levels and child literacy scores.

**Comparison of findings for teacher education.** In considering the range of programs, methods, and instruments in studies of preschool teacher educational levels, reasons for differential findings become apparent. Program types included Head Start, private, and state-funded preschools. Methodologies included regression, HLM, and growth modeling. Literacy measures included the ARS (West et al., 2000); Bayley II (Bayley, 1993); WJ-R (Woodcock & Johnson, 1990) and WJ-III (Woodcock et al., 2001) Letter Word Identification, Sound Awareness, and Incomplete Words subtests; and an unstandardized NECDL (2001) letter identification measure. The measures analyzed varied for some of the same studies, depending on the research teams (e.g., Early et al., 2007; NICHD ECCRN & Duncan, 2003). With such variability across studies, it is difficult to have consistency of results. Thus, a consistent linear association between teacher educational levels and child literacy outcomes was not established.
Teacher certification, years of experience, professional development, race and ethnicity. In addition to teacher educational levels, some preschool studies examined teacher characteristics including certification, years of experience, PD, and race and ethnicity (Early et al., 2007; Guo et al., 2010; Howes et al., 2008; Mashburn et al., 2006; NICHD ECCRN & Duncan, 2003; Pianta et al., 2005). Teaching certificates in preschool varied across and within preschool programs (i.e., Head Start, private and state-funded preschool) and included CDA, early childhood and elementary certifications, and special education endorsements (Early et al., 2007; Pianta et al., 2005). Years of experience was a numerical, continuous variable. PD included pre- and in-service training, workshops, and college courses for teaching staff (Pianta et al., 2005). Race and ethnicity was often defined as a dichotomous variable (White or non-White) or by multiple categories including African American, Asian American, Caucasian American, Hispanic American, Mixed race, or other (Mashburn et al., 2006, 2008; Pianta et al., 2005). No known studies to date examined linear associations between teacher certification, years’ experience, PD, race and ethnicity, and child literacy outcomes. In addition to the scarcity of studies of teacher demographic characteristics and child literacy outcomes, few preschool studies examined associations between teacher psychological characteristics and child literacy outcomes.

Psychological Teacher Characteristics and Child Literacy Outcomes

Preschool studies investigated teacher psychological characteristics including self-reported depression, child- or adult-centered beliefs, emotional supportiveness, and self-efficacy (Burchinal et al., 2008; Clarke-Stewart et al., 2002; Howes et al., 2008; Mashburn et al., 2006, 2008; Peisner-Feinberg et al., 2001; Pianta et al., 2005). Studies
reported mixed findings related to linear associations between teacher psychological characteristics and child literacy outcomes.

**Self-reported depression, child- or adult-centered beliefs.** While several studies examined preschool teacher self-reported depression (Clarke-Stewart et al., 2002; Mashburn et al., 2006; Pianta et al., 2005), none examined its linear association with child literacy outcomes. Similarly, there was interest in the examination of child- or adult-centered beliefs as one component of teacher attitudes that may be associated with improved child outcomes (Clarke-Stewart et al., 2002; Pianta et al., 2005); however, no studies examined their linear association with child literacy outcomes.

**Emotional supportiveness.** Preschool studies used various instruments to measure teacher emotional supportiveness and child literacy outcomes, and used differing methodologies to examine some of the same large datasets (Burchinal et al., 2008; Howes et al., 2008; Mashburn et al., 2008; Peisner-Feinberg et al., 2001). To measure preschool teacher emotional supportiveness, studies used one or both of the following instruments: the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008) and the Student-Teacher Relationship Scale (STRS; Pianta, 1992). The CLASS (Pianta et al., 2008) is a standardized measure that examines three components of teacher emotional supportiveness, including Classroom Climate, Teacher Sensitivity, and Regard for Student Perspectives. The STRS (Pianta, 1992) includes 30 items rated with 5-point Likert-type response options, ranging from 1 (definitely does not apply) to 5 (definitely applies); one of the three factors is Closeness. Three of the four preschool studies (Burchinal et al., 2008; Howes et al., 2008; Mashburn et al., 2008) analyzed CLASS data from the NCEDL studies. One of those studies (Howes et al., 2008) used both CLASS and STRS data. One additional study, the Cost, Quality, and Child Outcomes in Child
Care Centers (CQO; Peisner-Feinberg et al., 2001) used the STRS to examine teacher closeness. Peisner-Feinberg et al. (2001) used HLM analysis of CQO data to investigate childcare quality and children’s outcomes from preschool through second grade ($J = 167$ classrooms; $n = 733$ children) in four states estimated to be representative of childcare in the United States.


Further differences existed in analyses used by the four preschool studies (Burchinal et al., 2008; Howes et al., 2008; Mashburn et al., 2008; Peisner-Feinberg et al., 2001). Burchinal et al. (2008) conducted multiple regression analyses of Multi-State study data, whereas Howes et al. (2008) and Mashburn et al. (2008) used HLM to examine Multi-State and SWEEP study data. Peisner-Feinberg et al. (2001) conducted HLM analyses of CQO data.
Mixed evidence for emotional supportiveness. Of the four preschool studies that examined this teacher characteristic, only one found evidence for linear associations between teacher emotional supportiveness and child literacy outcomes (Howes et al., 2008). Howes et al. (2008) examined NCEDL Multi-State and SWEEP study CLASS (Pianta et al., 2008) and STRS (Pianta, 1992) data, and found evidence for linear associations between teacher closeness and child literacy outcomes, including 1) child letter naming ($\beta = .48, p < .05$), and 2) teacher-rated child literacy ($\beta = .17, p < .01$).

Using varying measures and methodologies as indicated, three of the four studies did not find evidence of linear associations between teacher emotional supportiveness and child literacy scores (Burchinal et al., 2008; Mashburn et al., 2008; Peisner-Feinberg et al, 2001). Mashburn et al. (2008) examined Multi-State and SWEEP study data, using CLASS but not STRS data. Burchinal et al. (2008) examined Multi-State study CLASS data, whereas Peisner-Feinberg et al. (2001) analyzed CQO study STRS data. None of the three studies found linear associations between teacher emotional supportiveness and child literacy scores.

Comparison of findings for emotional supportiveness. It is not surprising that results differed, as studies utilized different measures for teacher supportiveness (e.g., the CLASS, STRS) and for child literacy (e.g., the ARS [West et al., 2000], NCEDL [2001] letter identification measure, WJ-III [Woodcock et al., 2001] and WJ-R [Woodcock & Johnson, 1990] subtests, CTOPP [Wagner et al., 1999] subtest). Further, the studies used different methods of analysis, including regression and HLM. As a result, only one in four preschool studies found significant linear associations between teacher emotional supportiveness and child literacy outcomes (Howes et al., 2008).
Teacher self-efficacy. This section of the literature review presents studies with older children that established the significance of linear associations between TSE and child literacy outcomes (Armor et al., 1976; Ashton et al., 1982; Berman et al., 1977), and then introduces the one known preschool study that examined TSE and child literacy outcomes (Guo et al., 2010).

Positive evidence for older children. Studies with older children found that higher teacher self-efficacy was associated with increased children’s reading achievement (Armor et al., 1976; Ashton et al., 1982; Berman et al., 1977). Armor et al. (1976) conducted regression analyses of teacher Rand item scores and reading scores for urban Los Angeles elementary children ($n = 400$), and found that higher teacher efficacy was associated with increased child reading achievement. Berman and colleagues (1977) conducted correlational, factor analytic, and regression analyses of Rand item scores for elementary and secondary teachers ($n = 100$), and found positive linear associations between TSE and elementary children’s academic gains. Ashton et al. (1982) conducted a study with urban middle and high school teachers ($J = 97$) using correlation and regression analyses to examine the relationship between TSE and child academic outcomes. They found that teacher personal efficacy explained 46% of child reading score variance. Subsequent to these studies with teachers of older children, one preschool study found positive linear associations between TSE and child literacy outcomes (Guo et al., 2010).

Positive evidence for preschool children. One study examined linear associations between preschool TSE and child literacy outcomes. Guo et al. (2010) conducted correlational and HLM analyses of a subset ($n = 67$ teachers) of a larger language and literacy study to examine linear associations between teacher characteristics
and child literacy outcomes in selected Head Start, state-funded, and private preschools. They used an 11-item version of Bandura’s Teacher Efficacy Scale (hereinafter called Bandura’s TES; Bandura, 1997) and three measures of literacy: the Preschool Word and Print Awareness (PWPA; Justice & Ezell, 2001), and the Alphabet Knowledge and Name-Writing subtests of the Phonological Awareness and Literacy Screening-PreK (PALS; Invernizzi, Meier, & Sullivan, 2004). Guo et al. (2010) found a linear association between TSE and improved child print awareness, $\gamma_{01} = 0.022$, $t(63) = 3.45$, $p = .001$. To understand the significance of findings on linear associations between preschool TSE and child literacy outcomes, it is important to review the history of the measurement of self-efficacy.

**Measurement of Teacher Self-Efficacy**

Measures of TSE have evolved over the past 40 years and provide context for the current study’s psychometric analyses of the TSES. Theoretical views of TSE influenced the design and construction of measurement instruments, and generally reflected one of two approaches. The earliest instruments (e.g., from 1976-1982) developed according to the Rand items (Armor et al., 1976), and emphasized teacher beliefs in internal or external control of reinforcement, whereas later measures (e.g., from 1984-2001) followed Bandura’s (1986, 1997) approach of examining teachers’ beliefs about their capability to perform specific teaching tasks.

**Rand items’ approach and related measures.** Researchers sponsored by the Rand Corporation developed the first self-efficacy items based on Rotter’s (1966) social learning theory and beliefs about teacher efficacy (Armor et al., 1976). Rotter described a range of teacher beliefs about child learning. Some teachers believe that external, environmental factors (e.g., home and community values, violence, emphasis on
education) are greater than their influence on the learning process, whereas other teachers are confident in their abilities to influence child learning even in challenging situations (e.g., overcoming child learning difficulties, child achievement levels). The Rand items (Armor et al., 1976) used 5-point Likert-type response options ranging from 1 (strongly agree) to 5 (strongly disagree) and examined the extent to which teachers believed child learning was externally or internally controlled. The first Rand item reflected a teacher’s belief in his/her ability to influence factors from a child’s home environment (“When it comes right down to it, a teacher really can’t do much because most of a student’s motivation and performance depends on his or her home environment” [Berman et al., 1977, p. 159]). The second Rand item measured the extent to which a teacher could reach challenging children (“If I try really hard, I can get through to even the most difficult or unmotivated students” [Berman et al., 1977, p. 160]). Scoring included summing the two items, termed teacher efficacy (TE). These early studies did not report reliability or validity (Ashton et al., 1982; Berman et al., 1977).

Four subsequent TSE instruments (Ashton et al., 1982; Guskey, 1981, Rose & Medway, 1981,) expanded the Rand items (Armor et al., 1976) by adding items and response options that examined teachers’ beliefs about their ability to manage reinforcement and learning in the classroom. Two measures were published in 1981, the Teacher Locus of Control (TLC; Rose & Medway, 1981) and the Responsibility for Student Achievement (RSA; Guskey, 1981). The 28-item TLC assessed teachers’ attributions of child behaviors to teacher (e.g., internal) or environmental (e.g., external) sources. The items included forced-choice options, i.e., “Suppose you are teaching a student a particular concept in arithmetic or math and the student has trouble learning it. Would this happen (a) because the student wasn’t able to understand it, or (b) because
you couldn’t explain it very well?” (Rose & Medway, 1981, p. 189). TLC scores were weakly, positively, and significantly correlated with the Rand items ($r = .11$ to $.41$).

In the same year, Guskey (1981) expanded measurement of TSE by developing the RSA. The 30-item RSA was scored by teachers’ assigning a percentage of responsibility for measurement of positive and negative child outcomes (i.e., “If a student does well in your class, would it probably be [a] because that student had the natural ability to do well, or [b] because of the encouragement you offered?” [Guskey, 1981, p. 46]). Correlations between RSA and Rand items ranged from 0 for negative child outcomes to .81 for positive child outcomes; teachers were more confident in their abilities to effect positive outcomes than to prevent negative ones (Guskey, 1981). Soon after release of the TLC (Rose & Medway, 1981) and RSA (Guskey, 1981), National Institute of Education researchers developed two scales based on each of the Rand items (Armor et al., 1976).

Expanding measurement of the Rand items (Armor et al., 1976), Ashton et al. (1982) developed the Webb Efficacy Scale and the Ashton Vignettes. The Webb Efficacy Scale (Ashton et al., 1982) included 7 items with forced-choice type response options that examined General Teacher Efficacy. Teachers selected the option with which they agreed most strongly. For example, “A teacher should not be expected to reach every child; some students are not going to make academic progress” or “Every child is reachable; it is a teacher’s obligation to see to it that every child makes academic progress” (as described in Tschannen-Moran et al., 1998, p. 208). The Ashton Vignettes (Ashton et al., 1982) included 50 questions with forced-item response options that described Personal Teacher Efficacy dilemmas including motivation, discipline, instruction, planning, evaluation, and working with parents. As one example, “Your
school district has adopted a self-paced instructional program for remedial students in your area. How effective would you be in keeping a group of remedial students on task and engaged in meaningful learning while using these materials?” (as described in Tschannen-Moran et al., 1998, p. 209). Each of these aforementioned scales expanded measurement according to the Rand items (Armor et al., 1976); however, subsequent studies used them infrequently. In contrast, Bandura (1986, 1997) introduced a different approach to the measurement of self-efficacy that has been widely adopted, and that examined teacher beliefs in specific instructional areas.

**Bandura’s approach and related measures.** Self-efficacy scales developed since 1984 (Bandura, 1997; Gibson & Dembo, 1984; Tschannen-Moran & Woolfolk Hoy, 2001) followed Bandura’s (1986, 1997) approach for assessing teachers’ beliefs in their ability to teach specific tasks. Gibson and Dembo (1984) produced the Teacher Efficacy Scale (hereafter called Gibson & Dembo’s TES), a 30-item scale with two factors. Teaching Efficacy relates to a teacher’s general expectations for children’s outcomes, whereas Personal Teaching Efficacy indicates a teacher’s belief in his/her own ability to guide children’s learning. Orthogonal factor structure analyses resulted in factor pattern loadings ($\lambda$) that ranged from .45 to .65 for Teaching Efficacy, and from .46 to .61 for Personal Efficacy. Reported sample internal consistency of reliabilities ($\alpha$) ranged from .56 to .75 for Teaching Efficacy, and from .76 to .78 for Personal Efficacy (Allinder, 1994; Gibson & Dembo, 1984). Allinder’s (1994) sample included Midwestern special education elementary teachers ($n = 437$). While Gibson and Dembo’s TES factors followed Bandura’s (1997) approach, Bandura later articulated a seven-factor model of TSE in an instrument herein referred to as Bandura’s TES.
Bandura’s TES (Bandura, 1997) included 30 items with 9-point Likert-type response options, ranging from 1 (nothing) to 9 (a great deal). Bandura’s TES identified seven factors including Influencing Decision-Making and School Resources, Instructional and Disciplinary Efficacy, Enlisting Parent and Community Involvement, and Creating a Positive School Climate. Reliability and validity data were not reported (Bandura, 1997). NCEDL studies (Mashburn et al., 2006) and Guo et al. (2010) used abbreviated (10- and 11-item) versions of Bandura’s TES (Bandura, 1997). Guo et al.’s version further adapted the items with 5-point response Likert-type response options, ranging from 1 (nothing) to 5 (a great deal). The shortened versions included items such as “How much can you do to get through to the most difficult students?” and “How much can you do to keep students on task on difficult assignments?” (Guo et al., 2008, p. 1097). Reported reliabilities ranged from .85 to .90; no validity data were reported (Guo et al., 2010; Mashburn et al., 2006).

Based on an extensive review of self-efficacy measures, Tschannen-Moran and Woolfolk Hoy (2001) developed the TSES. The TSES includes 24 items; the original CFA analyses indicated three factors related to TSE: Instruction, Management, and Engagement, as well as one overall factor (Tschannen-Moran & Woolfolk Hoy, 2001). Interestingly, the correlations among factors ranged from .58 to .70 ($p < .001$), which suggested evidence for a general dimension present in each of the TSES factors. Further evidence for unidimensionality of the TSES was provided by Khairani and Razak (2012), in their investigation of the measurement properties of TSES scores with Malaysian pre-service and in-service teachers ($n = 213$). They conducted Rasch analyses with a modified (response options ranging from strongly agree to strongly disagree with an unspecified number of categories) version (Khairani & Razak, 2012).
Each TSES item includes a common stem: “How much can you do to…” or “How well can you…” or “To what extent can you…” and a specific teacher behavior (Tschannen-Moran & Woolfolk Hoy, 2001, p. 800). Likert-type response options range from 1 (nothing) to 9 (a great deal). TSES scores were positively correlated to the Rand items (Armor et al., 1976) ($r = .53$ and $.18$, $p < .01$) and to Bandura’s TES (1997) factors ($r = .16$, $p < .01$ for Teaching; $r = .64$, $p < .01$ for Personal Efficacy). Reliabilities for the total score ranged from .83 to .94 (Brown, 2005; Tschannen-Moran & Woolfolk Hoy, 2001). The current literature review revealed a need for psychometric examination of the TSES with preschool samples.

**Gaps in TSE Measurement.** Psychometric examination of preschool TSE, and especially of the TSES (Brown, 2005; Guo et al., 2010; Khairani & Razak, 2012), was limited and included factor (Long, 1983), correlational (Spearman, 1904), Rasch measurement model (Wright & Masters, 1982), and reliability analyses (Cronbach, 1951). Brown (2005) and Guo et al. (2010) used factor analyses (Long, 1983) within a CFA framework. Brown (2005) conducted a correlational study ($J = 94$) of teacher self-efficacy, beliefs about math instruction, and math instructional practices in one urban southeastern district, using the TSES. Khairani and Razak (2012) used Rasch measurement modeling to examine the Malaysian version of the TSES. They showed the TSES fit the unidimensional Rasch model and obtained linear measures from categorical data by transforming test scores into interval-scaled scores calibrated in log-odd units (Wright & Masters, 1982). Brown and Tschannen-Moran and Woolfolk Hoy (2001) reported reliability data to indicate the degree of internal consistency of TSES items. Tschannen-Moran and Woolfolk Hoy (2001) recommended additional psychometric examination of the TSES with other samples.
In addition to the analyses conducted to date for the TSES, psychometric theory (Crocker & Algina, 1986; Lord & Novick, 1968; Vandenberg & Lance, 2000) recommends additional steps including bifactor modeling (Reise et al., 2007) and longitudinal measurement invariance (LMI; Horn & McArdle, 1992; Jöreskog, 1971; Marsh, 1994). Specifically, modern CFA techniques examine the internal structure of measurement tools including one-factor, multi-factor, and bifactor structures (DeMars, 2013; Reise, 2012; Reise et al., 2007, 2010) when instruments consist of multiple correlated factors and when instruments are used to construct total and subscale scores. A one-factor model suggests that all item responses reflect one factor; a correlated-factors model suggests that multiple correlated factors reflect the content within the set of items. A bifactor model posits one overall or general factor, with specific factors that add unique, orthogonal, construct information (Quinn, 2014; Reise et al., 2007, 2010; Toland, Sulis, Giambona, Porcu, & Campbell, 2016). Holzinger and colleagues (Holzinger & Harman, 1938; Holzinger & Swineford, 1937) first introduced the bifactor model, in which both general and specific factors have direct and separate influence on items. Bifactor structures allow for more accurate examination of latent constructs than do unidimensional structures (Reise et al., 2007), and thus were included in analyses. In addition to bifactor modeling, psychometric theory recommended longitudinal measurement invariance (LMI; Horn & McArdle, 1992) of instruments, as described in the Psychometric Framework section, which was also included in current study analyses.

Statistical theorists recommend longitudinal growth modeling (LGM; McArdle, 1988; Meredith & Tisak, 1990) and HLM (Raudenbush & Bryk, 1986) analyses of instruments; see the Statistical Framework section for a description of each of these methods. While one study conducted HLM analyses of Bandura’s (1997) TES and child
literacy outcomes (Guo et al., 2010), no known studies had done so with the TSES. Therefore, each of these modeling steps was included in the current study, including bifactor modeling (Long, 1983), longitudinal measurement invariance (Horn & McArdle, 1992) and growth analyses (McArdle, 1988; Meredith & Tisak, 1990) of the TSES, as well as HLM (Raudenbush & Bryk, 1986) of teacher characteristics and child literacy outcomes. Further psychometric study with the TSES added to the knowledge base for preschool TSE, its measurement, and its association with child literacy outcomes.

The current study examined the measurement properties of the 24-item TSES and its relationship with child literacy outcomes in sample of children ages 3 to 6 years who attended state-funded preschool programs. Based on the literature review, the following research questions were tested:

1. Is there support for a one-, three-, or bifactor structure for the TSES?
2. Is there evidence for longitudinal measurement invariance (LMI) and longitudinal growth of TSES scores in the sample?
3. Do teacher characteristics (e.g., education, certification, years of experience, professional development, race and ethnicity, TSES score) influence growth in child literacy scores during preschool?

Based on the extant literature, the following hypotheses were developed:

1. A unidimensional model best represents the 24-item TSES.
2. Preschool teacher characteristics such as educational levels and TSES scores relate positively to growth in child literacy scores during one academic year.
Chapter 3: Method

Participants and Setting

A targeted sample of state-funded preschool teachers in Kentucky participated in this study. The University of Kentucky Office of Research Integrity provided Institutional Review Board approval of this exempt study as provided in Appendix A. The majority of districts (70%; 121 out of 173) used GOLD to assess child progress each year. All teachers in districts that used GOLD received invitations to participate. In the fall of 2013, of the 671 eligible teachers in the state, 273 (41%) responded and 265 (98%) completed all 24 TSES items. As summarized in Table 1, most teachers were female (95%), had a Master’s degree (64%), were certified in Interdisciplinary Early Childhood Education (IECE; 75%), had an average of 11 years of experience, had at least 10 hours of professional development (83%), and identified as White (94%). School characteristics included more teachers located in elementary schools (59%) than in blended Head Start programs (i.e., classrooms with state and Head Start funding, 22%), early childhood centers (15%), or Head Start centers (5%). Lastly, there were more teachers from rural (72%) than urban schools. In winter 2013, 196 teachers completed TSES items; and in spring 2014, 298 teachers completed TSES items; the demographics were similar across waves (see Table 1).
Table 1

Teacher Demographics (Percentages or SD in Parentheses) Each Wave Data Collection

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Fall (n = 265)</th>
<th>Winter (n = 196)</th>
<th>Spring (n = 298)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender^a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>254 (95.8)</td>
<td>185 (94.4)</td>
<td>296 (99.3)</td>
</tr>
<tr>
<td>Male</td>
<td>3 (1.1)</td>
<td>2 (1)</td>
<td>2 (0.7)</td>
</tr>
<tr>
<td>Highest degree</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GED</td>
<td>14 (5.3)</td>
<td>6 (3.1)</td>
<td>9 (3.0)</td>
</tr>
<tr>
<td>Associate’s</td>
<td>21 (7.9)</td>
<td>23 (11.7)</td>
<td>21 (7.0)</td>
</tr>
<tr>
<td>Bachelor’s</td>
<td>60 (22.6)</td>
<td>42 (21.4)</td>
<td>76 (25.5)</td>
</tr>
<tr>
<td>Master’s</td>
<td>170 (64.2)</td>
<td>125 (63.8)</td>
<td>192 (64.4)</td>
</tr>
<tr>
<td>Certification^b</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IECE</td>
<td>198 (74.7)</td>
<td>149 (76)</td>
<td>235 (78.9)</td>
</tr>
<tr>
<td>Elementary</td>
<td>95 (35.8)</td>
<td>69 (35.2)</td>
<td>100 (33.6)</td>
</tr>
<tr>
<td>CDA</td>
<td>52 (19.6)</td>
<td>45 (23)</td>
<td>53 (17.8)</td>
</tr>
<tr>
<td>Special education</td>
<td>51 (19.2)</td>
<td>35 (17.9)</td>
<td>54 (18.1)</td>
</tr>
<tr>
<td>Years Teaching^a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>30</td>
<td>40</td>
<td>41</td>
</tr>
<tr>
<td>Range</td>
<td>30</td>
<td>40</td>
<td>41</td>
</tr>
<tr>
<td>Mean</td>
<td>10.78 (7.16)</td>
<td>11.04 (7.59)</td>
<td>11.42 (7.85)</td>
</tr>
<tr>
<td>PD hours</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 hour</td>
<td>2 (0.8)</td>
<td>4 (2.0)</td>
<td>2 (0.7)</td>
</tr>
<tr>
<td>2-5 hours</td>
<td>18 (6.8)</td>
<td>6 (3.1)</td>
<td>12 (4.0)</td>
</tr>
<tr>
<td>6-10 hours</td>
<td>25 (9.4)</td>
<td>23 (11.7)</td>
<td>15 (5.0)</td>
</tr>
<tr>
<td>More than 10 hours</td>
<td>220 (83)</td>
<td>162 (82.7)</td>
<td>268 (89.9)</td>
</tr>
<tr>
<td>Race/Ethnicity^a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White, Non-Hispanic</td>
<td>249 (94)</td>
<td>190 (96.9)</td>
<td>275 (92.3)</td>
</tr>
<tr>
<td>African American</td>
<td>7 (2.6)</td>
<td>3 (1.5)</td>
<td>16 (5.4)</td>
</tr>
<tr>
<td>Hispanic American</td>
<td>2 (0.8)</td>
<td>1 (0.5)</td>
<td>1 (0.3)</td>
</tr>
<tr>
<td>Asian American</td>
<td>2 (0.8)</td>
<td>1 (0.5)</td>
<td>2 (0.7)</td>
</tr>
<tr>
<td>Other</td>
<td>4 (1.5)</td>
<td>1 (0.5)</td>
<td>2 (0.7)</td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary</td>
<td>(59)</td>
<td>(58)</td>
<td>(67)</td>
</tr>
<tr>
<td>Early childhood center</td>
<td>(15)</td>
<td>(11)</td>
<td>(13)</td>
</tr>
<tr>
<td>Blended Head Start</td>
<td>(22)</td>
<td>(27)</td>
<td>(21)</td>
</tr>
<tr>
<td>Head Start</td>
<td>(5)</td>
<td>(5)</td>
<td>(0)</td>
</tr>
<tr>
<td>Urbanicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>173 (72)</td>
<td>147 (75)</td>
<td>223 (75)</td>
</tr>
<tr>
<td>Urban</td>
<td>67 (28)</td>
<td>48 (25)</td>
<td>72 (24)</td>
</tr>
</tbody>
</table>

Note. IECE = Interdisciplinary Early Childhood Education; CDA = Child Development Associate; PD = Professional Development; GED = General Education Development test; ^aMissing values not included; ^bCategories not mutually exclusive.
The study utilized GOLD assessment data (approximately 20-40 children per teacher) for children of participating teachers, retrieved from the Kentucky Department of Education preschool assessment database. The database consisted of demographic characteristics including gender, race and ethnicity, special education services (through an Individual Educational Plan or IEP), age, and GOLD scores in 10 areas including literacy. As indicated in Table 2, the database included 8,996 children in fall; 6,018 children in winter; and 9,767 children in spring. In fall 2013, more children were boys (54%), White (80%), did not have an IEP (76%), and were an average of 53 months old; the sample had a mean literacy score of 29.98. Demographic statistics were similar across waves (see Table 2). As indicated in Table 2, sample statistics were compared to Kentucky and National Census Bureau (U. S. Census Bureau, 2015) rates for gender, race and ethnicity, and special education status. When compared to Kentucky’s children under the age of 5 years, sample percentages included more boys, fewer children who were White, and more children of mixed or other races. The sample included more children with disabilities, as state-funded preschool enrollment targets children who are at risk economically or by disability (Kentucky Governor's Early Childhood Task Force, 1999). Sample statistics differed from national statistics, including more children who were White and receiving special education services, and fewer children who were African American or Asian.
Table 2

Child Characteristics (Percent or SD in Parentheses) at Each Wave of Data Collection

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Fall ((n = 8,996))</th>
<th>Winter ((n = 6,018))</th>
<th>Spring ((n = 9,767))</th>
<th>KY Census 2015(^a)</th>
<th>U.S. Census 2015(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender(^b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td>4,836 (54)</td>
<td>3,240 (54)</td>
<td>5,306 (54)</td>
<td>(51)</td>
<td>(51)</td>
</tr>
<tr>
<td>Girls</td>
<td>4,128 (46)</td>
<td>2,757 (46)</td>
<td>4,425 (45)</td>
<td>(49)</td>
<td>(49)</td>
</tr>
<tr>
<td>Race and Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>7,162 (80)</td>
<td>4,754 (79)</td>
<td>7,766 (80)</td>
<td>(88)</td>
<td>(68)</td>
</tr>
<tr>
<td>American Indian</td>
<td>22 (0)</td>
<td>15 (0)</td>
<td>21 (0)</td>
<td>(0)</td>
<td>(1)</td>
</tr>
<tr>
<td>Asian</td>
<td>60 (1)</td>
<td>31 (1)</td>
<td>56 (1)</td>
<td>(1)</td>
<td>(5)</td>
</tr>
<tr>
<td>African</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American</td>
<td>769 (9)</td>
<td>529 (9)</td>
<td>856 (9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawaiian</td>
<td>17 (0)</td>
<td>15 (0)</td>
<td>19 (0)</td>
<td>(0)</td>
<td>(0)</td>
</tr>
<tr>
<td>More than one</td>
<td>500 (6)</td>
<td>377 (6)</td>
<td>546 (6)</td>
<td>(2)</td>
<td>(6)</td>
</tr>
<tr>
<td>Other</td>
<td>466 (5)</td>
<td>297 (5)</td>
<td>503 (5)</td>
<td>(1)</td>
<td>(6)</td>
</tr>
<tr>
<td>Special Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No IEP</td>
<td>6,810 (76)</td>
<td>4,528 (75)</td>
<td>6,926 (71)</td>
<td>(95)</td>
<td>(96)</td>
</tr>
<tr>
<td>With IEP</td>
<td>2,186 (24)</td>
<td>1,490 (25)</td>
<td>2,841 (29)</td>
<td>(5)</td>
<td>(4)</td>
</tr>
<tr>
<td>Age in months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>35</td>
<td>36</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>75</td>
<td>78</td>
<td>81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>41</td>
<td>43</td>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ((SD))</td>
<td>53 (5.99)</td>
<td>55 (6.69)</td>
<td>58 (7.07)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GOLD literacy score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>88</td>
<td>103</td>
<td>104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>88</td>
<td>103</td>
<td>104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ((SD))</td>
<td>29.98 (13.25)</td>
<td>39.23 (15.96)</td>
<td>51.29 (20.14)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. IEP = Individual Education Plan. \(^a\)Children under 5 years; \(^b\)Does not include missing data.

Instrumentation

Teachers’ Sense of Efficacy Scale (TSES; Tschannen-Moran & Woolfolk)

Hoy, 2001). Tschannen-Moran and Woolfolk Hoy (2001) developed the TSES, based on Bandura’s theory and considerable research of TSE measures. The TSES assesses three separate (Instruction, Management, Engagement) and one overall factor, through 24- and 12-item versions. The 24-item scale has 8 items for each factor; the 12-item version
includes 4 items for each factor. Each item is rated on a 9-point Likert type response option ranging from 1 (nothing) to 9 (a great deal). Average TSES mean item scores for the current sample ranged from 6.86 to 7.89 in fall, from 6.90 to 7.90 in winter, and from 7.01 to 8.08 in spring. The TSES also includes eight questions about teacher and school characteristics, including teachers’ years of experience, race and ethnicity, and gender; and school characteristics including economic and urban status, subject matter taught, grade, and school level. The TSES is copyrighted and available for research and non-profit purposes. The current study modified the demographic questions for a preschool sample based on areas of interest indicated in the review of literature, substituting educational level, school building location, and professional development hours for subject matter, grade, and school level. A copy of the TSES is provided in Appendix B.

Teaching Strategies GOLD (Heroman et al., 2010). GOLD is the most frequently used assessment in state-funded preschool programs across the nation (Ackerman & Coley, 2012; Lambert, Kim, & Burts, 2015). GOLD assesses children aged birth to six years in ten areas of development, including cognitive, language, literacy, math, social-emotional, physical, English language acquisition, science and technology, social studies, and arts; see Appendix C. GOLD includes 38 objectives for the ten areas; dimensions and examples define each objective. Twenty-five objectives including cognitive (4 objectives), language (3 objectives), literacy (5 objectives), math (4 objectives), social-emotional (3 objectives), physical (4 objectives), and English language acquisition (2 objectives) are rated on a 10-point scale from 0 (not yet) to 9 (kindergarten level). Approximate age intervals for these 25 objectives are included. Thirteen additional objectives for science and technology (5 objectives), social studies (4 objectives), and arts (4 objectives) are rated on a scale from 0 (no evidence) to 2 (meets
program expectations). Increased scores for each objective indicate child progress; results include scores for developmental areas, objectives, and dimensions. Lambert et al. (2015) reported internal consistency of reliability (α) for each developmental area ranging from .97 (social-emotional) to .98 (cognitive). Current study omega reliabilities for child literacy scores, based on the weighted least squares means and variance adjusted (WLSMV) estimator for categorical data in Mplus, included .976 (95% bootstrap corrected CI [.974, .978]) for fall, .976 (95% bootstrap corrected CI [.971, .980]) for winter, and .977 (95% bootstrap corrected CI [.976, .978]) for spring.

Study Design

Data collection occurred during the 2013-2014 academic year. Kentucky preschool teachers completed the TSES three times (fall, winter, spring) to examine changes in TSE over time. GOLD child assessments occurred two to three times annually (fall, winter [optional], spring) according to state and district policy. Teachers received TSES surveys through a Qualtrics-developed secure survey. For fall and winter distributions, Preschool Coordinators received invitation emails in each district. Coordinators distributed the survey link and instructions to preschool teachers. The coordinator’s email to instructors is in Appendix D. In spring 2014, all eligible teachers received survey links directly via their school email addresses. Teachers had two weeks to complete the survey, with email reminders sent at weeks one and two. Teachers took from two to 41 minutes to complete the survey. Teachers who completed all three surveys were eligible for one of four gift cards, valued at $75 each.

Data Analyses

Data analyses included psychometric examination of TSES scores within a CFA framework (Long, 1983), which provided useful methods for examining factor structures,
measurement invariance across time (Jöreskog, 1971), and model-based reliability estimates. Statistical analyses included longitudinal growth modeling (McArdle, 1988; Meredith & Tisak, 1990) of TSES scores and HLM (Raudenbush & Bryk, 1986) of teacher characteristics and child literacy outcomes.

**Psychometric analysis of TSES scores.**

**Confirmatory factor analyses.** The first step of psychometric analysis included examination of the internal structure of the 12- and 24-item TSES within a CFA framework using Mplus version 7.4 (Muthén & Muthén, 1998-2015). Based on the identification of one-factor (Khairani & Razak, 2012) and one- and three-factor structures (Tschannen-Moran & Woolfolk Hoy, 2001), as well as strong correlations among factors (.58 to .70; Tschannen-Moran & Woolfolk Hoy, 2001), three models were examined, including one-factor (unidimensional), three-factor, and bifactor models. Reise et al. (2007, 2010) and Rodriguez et al. (Rodriguez, Reise, & Haviland, 2016a, 2016b) recommended that researchers complement analyses of the one-factor and correlated factors model with the bifactor model. Figure 1 presents the conceptual representation of these three models with respect to the 24-item TSES.
Figure 1. One-factor, correlated three-factor, and bifactor model conceptual representation of the 24-item TSES.

Due to the categorical nature of TSES item responses, it would be appropriate to conduct CFA analyses with the WLSMV estimator, which is, in essence, a CFA of the
polychoric correlations. However, this estimator employs pairwise deletion and does not allow use of the full sample size (as it assumes data are missing completely at random). Thus, using WLSMV may result in non-identified models due to its handling of missing data. Therefore, study analyses used the robust maximum likelihood (MLR) estimator, because it assumed data were missing at random, used all available data, and was appropriate for categorical data, especially with the 9-category response option format of the TSES (Bandalos & Finney, 2001; Rubin, 1976).

Several indices were used to evaluate the fit of the CFAs, as no single index was perfect and each index gave different and important information for consideration. Indices included the global robust chi-square ($\chi^2$) index (which indicated a lack of fit from over-identification of restrictions on the model or from smaller sample sizes) via MLR and its associated $p$ value. Since use of the chi-square index alone was an identified concern (Brown, 2006), additional model fit indices were considered. Additional indices included the root mean square error of approximation (RMSEA; Steiger, 1989; an absolute index that examined closeness of fit); the standardized root mean square residual (SRMR; Bentler, 2006; Hu & Bentler, 1999; a measure of standardized difference between observed and predicted correlations), and the comparative fit index (CFI; Bentler, 1990; an index that accounted for smaller sample sizes). Recommended criteria for each index included lower $\chi^2$ values with non-significant (> .05) $p$ values; RMSEA values less than .08; SRMR values less than .10 (Hu & Bentler, 1999; Weston & Gore, 2006); and CFI values greater than .90 (Medsker, Williams, & Holahan, 1994). These model-fit criteria may not be directly applicable to categorical data, as they were primarily derived using maximum likelihood estimation for continuous indicators.
In addition, information indices designed for use with non-nested models (Burnham & Anderson, 2002), such as the correlated factors and bifactor models in this study, compared non-nested and nested CFA models. Specifically, the Akaike information criterion (AIC; Akaike, 1987) and Bayesian information criterion (BIC; Schwarz, 1978) indices were used to make relative model-data fit comparisons. AIC and BIC indices were not used in isolation, but rather for comparing competing models that were non-nested or nested. Lower values for AIC and BIC indicated better fit (Hu & Bentler, 1999; van de Shoot et al., 2012). In addition, AIC and BIC value differences that were greater than 6, and especially greater than 10, provided evidence of acceptable and strong model fit difference, respectively (Burnham & Anderson, 2002; Fabozzi, Focardi, Rachev, & Arshanapalli, 2014; Kass & Raftery, 1995). A nominal .05 significance level was applied to all statistical analyses, unless indicated otherwise. In the current study, models that did not demonstrate adequate fit were not included in subsequent analyses.

Consistent with recommendations by Reise et al. (2007, 2010), analyses included two additional measures of dimensionality from the bifactor solution: explained common variance (ECV; Reise et al., 2010; ten Berge & Sočan, 2004) and percent of uncontaminated correlations (PUC; Reise, Scheines, Widaman, & Haviland, 2013). ECV estimated the amount of variance that was due to the general factor for a scale. Review of the bifactor literature within the CFA framework indicated guidelines for ECV and PUC, as well as recommended relationships between the indices. Stucky, Thissen, and Edelen (2013) suggested that an ECV value greater than .85 for the general factor represented a unidimensional scale, whereas Quinn (2014) suggested a cutoff of .90 for binary data. Reise et al., 2013 recommended PUC values greater than .80, as an
important index for moderating the effects of imposing a unidimensional structure on multidimensional measures. Further guidelines included several recommendations for combined values of ECV and PUC indices. When PUC and ECV values were at least .70, or when ECV values were greater than .60 and PUC values less than .80, relative bias would be slight and the measure could be considered essentially unidimensional. However, if PUC and ECV values were both less than .60, a unidimensional solution could result in biased parameter estimates (Reise et al., 2013; Rodriguez et al., 2016a).

In addition to examination of ECV and PUC values, analyses included inspection of the correlations between factors from the correlated factors model. Correlations between factors were another indicator of instrument dimensionality, as high correlations may have indicated redundancy and potential dependence among items (Quinn, 2014; Reise et al., 2013).

**Longitudinal measurement invariance.** LMI (Jöreskog, 1971) tests were the second step of psychometric examination, using a CFA approach as recommended by Coertjens (2014) and van de Shoot et al. (2012). Three increasingly restrictive models (baseline, metric, scalar) examined LMI, to determine whether the latent construct of TSE had the same meaning across the three waves of data. First, the baseline (or configural) model “Constrained items to associate with the same factors across waves, but factor loadings, unique error variances, and item intercepts were freely estimated” (Buhs, McGinley, & Toland, 2010, p. 181) across waves of data collection. The first model examined whether TSES factor structure was the same across waves, with items associating with the same factors. Second, the metric (or weak factorial) model constrained factor loadings across the waves of data collection, with freely varying error variance and item intercepts. The second model examined whether model fit was
achieved when factor loadings were held equal across waves of data. Third, for the scalar (or strong) model, factor loadings and item intercepts were constrained across the waves of data collection while item error variance was freely estimated, to determine whether model fit was achieved when factor loadings and item intercepts were the same across waves of data (Buhs et al., 2010; van de Shoot et al., 2012).

Model fit indices included $\chi^2$ index and its associated $p$ value, RMSEA, SRMR, and CFI. Relative fit between models (i.e., comparisons between the three models) included examination of change in $\chi^2$ ($\Delta \chi^2_{MLR}$); recommended criteria stated that $\Delta \chi^2$ with $p$ values greater than .01 suggest noninvariance for the measurement invariance testing process (Byrne, 2010). The chi-square difference test for nested models examined significance differences between each successive model, by comparing the more restricted model (e.g., metric invariance) with the less restricted model (e.g., baseline invariance). When no significant difference was found, the more restrictive model was accepted. If a significant difference was found, the less restrictive model was deemed tenable (Buhs et al., 2010).

Relative fit between models was also examined with change in CFI ($\Delta$CFI); criteria included $\Delta$CFI less than or equal to -.01 indicating acceptable invariance (Cheung & Rensvold, 2002). Relative model fit indices also included AIC and BIC differences as indicated, with differences between models greater than 6 providing evidence of acceptable model-data fit improvement (Burnham & Anderson, 2002; Fabozzi, Focardi, Rachev, & Arshanapalli, 2014; Kass & Raftery, 1995). Difference values less than 6 were considered evidence of invariance occurring between competing models.

**Reliability analyses.** Lastly, analyses examined omega reliability of TSES scores for the sample. Omega coefficients were appropriate to the current study, due to the
congeneric (Jöreskog, 1971; Lord & Novick, 1968) nature of TSES items (i.e., items measured the same common factor, though they may have had different factor loadings and error variances). The omega coefficient relaxed the assumption of equal factor loadings (based on the tau equivalent measurement model) and was therefore a more realistic estimate of the population reliability coefficient (Dunn, Baguley, & Brunsden, 2014; Geldhof, Preacher, & Zyphur, 2014; Revelle & Zinbarg, 2009; Sijtsma, 2009).

**Statistical analyses.**

*Longitudinal growth modeling.* To examine growth in TSES scores over time, longitudinal growth modeling (McArdle, 1988; Meredith & Tisak, 1990) was conducted. Both growth modeling (i.e., using item sum scores) and latent growth modeling (based on multiple indicators for the TSES) for each wave of data were conducted in Mplus version 7.4 (Muthén & Muthén, 1998-2015) for the 12- and 24-item TSES versions. For growth and latent growth modeling, Model 1 examined longitudinal growth on the TSES without predictors. Model 2 examined longitudinal growth on the TSES after accounting for teacher characteristics at the intercept (i.e., fall data), and Model 3 examined longitudinal growth on the TSES after accounting for teacher characteristics at the intercept and interactions between each teacher characteristic and time (i.e., wave of data). Model fit indices included $\chi^2_{MLR}$ and its associated $p$ value, RMSEA, CFI, and SRMR. Model comparison indices included $\Delta\chi^2$, change in $df$ ($\Delta df$), AIC, and BIC. Given the focus on improvement in model fit between each subsequent model (as in hierarchical regression when testing changes between $R^2$ change), interpretation of results emphasized model comparison and not model fit indices.

**HLM of teacher characteristics and child literacy scores.** HLM 7.01 for Windows (Raudenbush, Bryk, & Congdon, 2011) was used to examine linear
associations between growth in teacher characteristics (including TSES scores) and child literacy scores. A three-level HLM examined child literacy outcome scores while controlling for four variables at the child level (gender, race and ethnicity, special education status, age in fall) and seven variables at the teacher level (education, certification, years of experience, race and ethnicity, TSES fall score, classroom location, classroom urbanicity).

Due to limited variability, teacher gender (95% female) and professional development scores with values ranging from 0 (less than 1 hour) to 3 (10 hours or more) with $SD = .39$ were not included in HLM analyses. Moreover, due to the limited variability of teacher and child characteristics (see Table 3), non-dichotomous variables were dichotomized as had been done in prior studies (Early et al., 2007; Howes et al., 2008; Mashburn et al., 2006, 2008; Pianta et al., 2005). For child variables, race and ethnicity categories included White (80%) or non-White (20%). Child data collection intervals were coded according to chronological time in months for each wave, including 0 (fall), 2.53 (winter), and 6.26 (spring) months. For teacher variables, educational levels included having a Master’s degree (65%) or less than a Master’s (i.e., GED, AA, BA; 35%); certification included having an IECE (76%) or other certification (i.e., CDA, Elementary, Special Education; 24%). Race and ethnicity included White (95%) or non-White (5%); classroom location included public school (elementary building and early childhood center, 73%) or Head Start (Blended Head Start and Head Start classrooms, 27%). Classroom urbanicity included rural (72%) and urban (28%) classrooms. Table 3 presents descriptive data for children and teachers in HLM Levels 1, 2, and 3.
Table 3

*Descriptive Data for Children and Teachers (Percentages or SD in Parentheses): HLM*

**Levels 1, 2, and 3**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1- Child Literacy Over Time</td>
<td>N</td>
<td>M (SD)</td>
<td>Minimum</td>
<td>Maximum</td>
<td></td>
</tr>
<tr>
<td>Literacy growth</td>
<td>12,304</td>
<td>40.16 (18.95)</td>
<td>0</td>
<td>103</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characteristic</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 2- Child Characteristics</td>
<td>N (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td>5,289 (55%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls</td>
<td>4,385 (45%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>7,713 (80%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-White</td>
<td>1,961 (20%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Special Educ status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No IEP</td>
<td>6,970 (72%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With IEP</td>
<td>2,704 (28%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall age in months</td>
<td>9,674</td>
<td>52.27 (6.83)</td>
<td>30</td>
<td>75</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N (%)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 3- Teacher Characteristics</td>
<td>N (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master’s</td>
<td>157 (65%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>83 (35%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IECE</td>
<td>183 (76%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>57 (24%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>229 (95%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-White</td>
<td>11 (5%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary</td>
<td>176 (73%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head Start</td>
<td>64 (27%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urbanicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>173 (72%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>67 (28%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years teaching</td>
<td>240</td>
<td>11.28 (7.15)</td>
<td>0</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>PD hours</td>
<td>240</td>
<td>2.92 (.39)</td>
<td>0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Fall TSES</td>
<td>240</td>
<td>175.93 (21.99)</td>
<td>111</td>
<td>216</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* IEP = Individual Education Plan; TSES = Teachers’ Sense of Efficacy Scale; PD = Professional Development; IECE = Interdisciplinary Early Childhood Education.
HLM models were built as follows: in Level 1, the baseline model, child literacy outcome scores were entered \( n = 12,304 \) children, without any child- or teacher-level predictors. HLM permitted missing values at this level only. In Level 2, the four child demographic variables were entered \( n = 9,674 \) children with complete data), for both intercept and rate of growth. In Level 3, the seven teacher variables were entered \( n = 240 \) cases with complete data) for both intercept and rate of growth. Each HLM level was specified as follows.

The Level 1 model was situated at the child level and included child literacy score growth over time. The Level 1 regression equation modeled child literacy growth on time in preschool for each child as follows:

\[
LIT_{tij} = \pi_{0ij} + \pi_{1ij}(child\_time\ interval)_{tij} + rt_{ij} \tag{1}
\]

where \( LIT_{tij} \) is the literacy outcome growth score \( t \) for child \( i \) with teacher \( j \); \( \pi \) is the intercept value for literacy scores; \( child\_time\ interval \) includes the fall, winter, or spring data point; and \( r_{tij} \) is the error term that is unique at each time \( t \) for child \( i \) with teacher \( j \).

The Level 2 model was situated at the child level and included four child background covariates. It was modeled with two regression equations for between-child initial status and rate of growth as follows:

\[
\pi_{0ij} = \beta_{00j} + \beta_{10j}(child\_age)_{ij} + \beta_{20j}(child\_gender)_{ij} + \beta_{30j}(child\_race)_{ij} + \beta_{40j}(child\_special\ education)_{ij} + u_{0ij} \tag{2}
\]

\[
\pi_{1ij} = \beta_{10j} + \beta_{11j}(child\_age)_{ij} + \beta_{21j}(child\_gender)_{ij} + \beta_{31j}(child\_race)_{ij} + \beta_{41j}(child\_special\ education)_{ij} + u_{1ij} \tag{3}
\]

where \( \pi_{0ij} \) is the average initial literacy score for children \( i \) with teacher \( j \); \( \beta_{00j} \) is the intercept value for teacher \( j \); \( \beta_{10j}, \beta_{20j}, \beta_{30j}, \) and \( \beta_{40j} \) are regression coefficients for each of the child entry level variables (i.e., age, gender, race, special education status). Lastly,
is the child-level random error term for child \( i \) with teacher \( j \) at the teacher level. Further, \( \pi_{ij} \) is the average literacy growth score for children \( i \) with teacher \( j \); and \( \beta_{11j}, \beta_{21j}, \beta_{31j}, \) and \( \beta_{41j} \) are the regression coefficients for the child growth variables (age, gender, race, special education status), and \( u_{ij} \) is the child-level random error term for child \( i \) with teacher \( j \) at the teacher level. Child age was centered around the grand mean, so that \( \beta_{00j} \) and \( \beta_{01j} \) represent the initial status and rate of growth for the average age of preschool children in the sample.

The Level 3 model was situated at the teacher level and included seven teacher variables, including fall TSES scores for 265 cases (of which 240 cases had full data and 204 cases had matching child and teacher data). It included two between-teacher regression equations that modeled average initial status and average rate of growth in child literacy on seven teacher characteristics as follows:

\[
\beta_{00j} = \gamma_{000} + \gamma_{001}(education)_j + \gamma_{002}(certification)_j + \gamma_{003}(years\_teaching)_j + \\
\gamma_{004}(race)_j + \gamma_{005}(TSES)_j + \gamma_{006}(location)_j + \gamma_{007}(urbanicity)_j + v_{00j} \tag{4}
\]

\[
\beta_{01j} = \gamma_{010} + \gamma_{011}(education)_j + \gamma_{012}(certification)_j + \gamma_{013}(years\_teaching)_j + \\
\gamma_{014}(race)_j + \gamma_{015}(TSES)_j + \gamma_{016}(location)_j + \gamma_{017}(urbanicity)_j + v_{01j} \tag{5}
\]

where \( \beta_{00j} \) and \( \beta_{01j} \) represent average initial status and average rate of growth; \( \gamma_{000} \) and \( \gamma_{010} \) are initial status and rate of growth intercepts, respectively; \( \gamma_{001}-\gamma_{007} \) are regression coefficients for the seven teacher variables at initial status; \( \gamma_{011}-\gamma_{017} \) are regression coefficients for the seven teacher variable rates of growth. Lastly, \( v_{00j} \) and \( v_{01j} \) are random error terms for the intercept and slope at the teacher level. Continuous variables including teacher years of experience and TSES scores were centered around the grand mean. Analyses used full information maximum likelihood estimation with robust
standard errors, which used all available data. All statistical significance tests were performed at an alpha level of .05.

HLM analyses examined fixed effects, random effects, and the proportion of variance for each level of analysis, as well as variable effect size (Bryk & Raudenbush, 1992; Raudenbush & Bryk, 1986, 2002). Examination of fixed effects included calculating the average effect between predictors and the outcome variable (child literacy growth), at both child and teacher levels. Examination of random effects included consideration of the effects of intercepts varying by individuals and by classrooms/teachers, at both child and teacher levels. Analyses included examination of the variance proportioned at each level, for child and teacher level variables. The proportion of variance indicated the amount of variance in scores associated with each variable (Raudenbush & Bryk, 1988). Effect sizes were calculated by dividing the coefficients of variables at the intercept and growth rates in the final model by the variance (standard deviation) for the intercept and growth rate, respectively, in the baseline model. For example, the effect size for a variable at the intercept level in the final model was divided by the standard deviation for the intercept in the baseline model. Effect sizes were classified according to Rosenthal and Rosnow’s (1984) criteria as large (.50 or greater), moderate (.30 to .50), small (.10 to .30) or trivial (less than .10).
Chapter 4: Results

Results for Psychometric Analyses of TSES Scores

Confirmatory factor analyses.

Model fit results for the 24-item TSES. To address the first research question (TSES dimensionality), CFAs were conducted in *Mplus* 7.4 ([Muthén & Muthén](https://www.statmodel.com)), for fall, winter, and spring data for three models (unidimensional, three-factor, bifactor), for the 24-item (see Table 4) and 12-item (see Table 5) TSES. Beginning with the 24-item TSES, CFA model fit indices ($\chi^2_{\text{MLR}}$, RMSEA, CFI, SRMR) informed examination of each of the three models at each wave of data collection, as presented in Table 4.

For the fall data, as presented in Table 4, model fit indices showed satisfactory fit for the bifactor model, $\chi^2_{\text{MLR}}(228) = 443.23$, $p < .001$, CFI = .94, RMSEA = .06, SRMR = .04. Results indicated adequate fit for the 3-factor model, $\chi^2_{\text{MLR}}(249) = 500.62$, $p < .001$, CFI = .93, RMSEA = .06, SRMR = .04. Model fit indices showed reasonable fit for the 1-factor model, $\chi^2_{\text{MLR}}(252) = 625.88$, $p < .001$, CFI = .89, RMSEA = .08, SRMR = .05. Comparative model fit indices informed further comparison of models for the fall data, including AIC, BIC, and the chi-square difference test. AIC values showed the bifactor model was better fitting, whereas the BIC suggested the 3-factor model. The chi-square difference test showed the 3-factor model had better fit than the 1-factor model, $\chi^2_{\text{MLR}}(3) = 67.06$, $p < .001$.

For the winter data, as presented in Table 4, model fit indices showed satisfactory fit for the 3-factor model, $\chi^2_{\text{MLR}}(249) = 385.12$, $p < .001$, CFI = .95, RMSEA = .05, SRMR = .05. The bifactor model ($\chi^2_{\text{MLR}}[228] = 473.69$, $p < .001$, CFI = .88, RMSEA = .07, SRMR = .05) showed adequate fit; the 1-factor model ($\chi^2_{\text{MLR}}[252] = 481.59$, $p <$
.001, CFI = .91, RMSEA = .07, SRMR = .05) showed reasonable fit. Moreover, AIC and BIC indices favored the 3-factor model relative to the 1-factor and bifactor models. As indicated in Table 4, the chi-square difference test showed the 3-factor model had better fit than the 1-factor model, \( \chi^2_{\text{MLR}}(3) = 73.38, p < .001 \).

For the spring data, as shown in Table 4, model fit indices showed satisfactory fit for the 3-factor model, \( \chi^2_{\text{MLR}}(249) = 480.10, p < .001, \text{CFI} = .94, \text{RMSEA} = .06, \text{SRMR} = .04 \). Model fit indices showed adequate fit for the bifactor (\( \chi^2_{\text{MLR}}[228] = 574.26, p < .001, \text{CFI} = .82, \text{RMSEA} = .07, \text{SRMR} = .05 \)) model and reasonable fit for the 1-factor model, \( \chi^2_{\text{MLR}}(252) = 626.79, p < .001, \text{CFI} = .90, \text{RMSEA} = .07, \text{SRMR} = .05 \). Moreover, AIC and BIC indices favored the 3-factor model relative to the 1-factor and bifactor models. The chi-square difference test showed the 3-factor model had better fit than the 1-factor model, \( \chi^2_{\text{MLR}}(3) = 98.92, p < .001 \).

Overall, model fit results for the 24-item TSES showed that that the 3-factor and bifactor models fit the data best for fall, and the 3-factor model fit the data best in winter and spring. Although the results suggested that, overall, the 3-factor model was the best fit to the data across the three waves of data; the correlations among factors in the 3-factor model were high, suggesting redundancy in content. The correlations ranged from .88 to .95 for fall, from .86 to .97 for winter, and from .86 to .96 for spring data. Also, Reise et al. (2010) and Rodriguez et al. (2016a, 2016b) suggested that when a one-factor model was applied to multidimensional data and a general factor applied to all items, poor model fit would likely occur; however, bifactor results suggested that fit for a one dimensional model was reasonable.

To clarify the dimensionality of the 24-item TSES in this study, further analyses included ancillary measures based on the bifactor model solution as discussed in the Data
Analyses section, including ECV and PUC indices (Reise et al., 2013; Rodriguez et al., 2016b). ECV values were .89 in fall, .85 in winter, and .87 in spring for the general factor of the bifactor model, supporting a unidimensional model (Quinn, 2014; Stucky & Edelen, 2015). The PUC value for the TSES was .70. Based on Reise et al.’s (2013) criteria, PUC and ECV indices further supported a unidimensional structure. Relatively speaking, the 3-factor model fit was best, but ancillary bifactor solution results indicated sufficient evidence to use a one-factor model for multidimensional data, with unbiased parameters in the one-factor model.
### Table 4

**CFA Model Fit Results for 1-Factor (1-fac), 3-Factor (3-fac), and Bifactor (Bifac) Models to the 24-Item TSES using MLR Estimation**

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2_{MLR}$</th>
<th>$df$</th>
<th>CI [LL, UL]</th>
<th>CFI</th>
<th>SRMR</th>
<th>AIC</th>
<th>BIC</th>
<th>$\chi^2$ diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall (n = 265)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-fac</td>
<td>625.88</td>
<td>252</td>
<td>.08 [.07, .08]</td>
<td>.89</td>
<td>.05</td>
<td>1,6106.81</td>
<td>1,6364.55</td>
<td></td>
</tr>
<tr>
<td>3-fac</td>
<td>500.62</td>
<td>249</td>
<td>.06 [.05, .07]</td>
<td>.93</td>
<td>.04</td>
<td>15,933.77</td>
<td>16,202.24</td>
<td>67.06</td>
</tr>
<tr>
<td>Bifac</td>
<td>443.23</td>
<td>228</td>
<td>.06 [.05, .07]</td>
<td>.94</td>
<td>.04</td>
<td>15,871.83</td>
<td>16,215.48</td>
<td></td>
</tr>
<tr>
<td>Winter (n = 196)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-fac</td>
<td>481.59</td>
<td>252</td>
<td>.07 [.06, .08]</td>
<td>.91</td>
<td>.05</td>
<td>11,338.48</td>
<td>11,574.51</td>
<td></td>
</tr>
<tr>
<td>3-fac</td>
<td>385.12</td>
<td>249</td>
<td>.05 [.04, .06]</td>
<td>.95</td>
<td>.05</td>
<td>11,213.66</td>
<td>11,459.52</td>
<td>73.38</td>
</tr>
<tr>
<td>Bifac</td>
<td>473.69</td>
<td>228</td>
<td>.07 [.07, .08]</td>
<td>.88</td>
<td>.05</td>
<td>12,473.38</td>
<td>12,788.08</td>
<td></td>
</tr>
<tr>
<td>Spring (n = 298)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-fac</td>
<td>626.79</td>
<td>252</td>
<td>.07 [.06, .08]</td>
<td>.90</td>
<td>.05</td>
<td>16,758.62</td>
<td>17,024.81</td>
<td></td>
</tr>
<tr>
<td>3-fac</td>
<td>480.10</td>
<td>249</td>
<td>.06 [.05, .06]</td>
<td>.94</td>
<td>.04</td>
<td>16,551.54</td>
<td>16,828.82</td>
<td>98.92</td>
</tr>
<tr>
<td>Bifac</td>
<td>574.26</td>
<td>228</td>
<td>.07 [.06, .08]</td>
<td>.82</td>
<td>.05</td>
<td>20,013.87</td>
<td>20,368.79</td>
<td></td>
</tr>
</tbody>
</table>

*Note. TSES = Teachers’ Sense of Efficacy Scale; 1-fac = 1-factor; 3-fac = 3-factor; Bifac = Bifactor; MLR = robust maximum likelihood; $df$ = degrees of freedom; RMSEA = root mean square error of approximation; CFI = comparative fit index; SRMR = standardized root mean square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion; $\chi^2$ diff = chi-square difference. All models were statistically significant at $p < .001$.\n
**Model fit results for the 12-item TSES.** To explore further the dimensionality of the TSES, CFA modeling was conducted for the 12-item TSES. CFA model fit indices ($\chi^2_{MLR}$, RMSEA, CFI, SRMR) informed examination of the three models at each wave of data, as presented in Table 5. Results were similar to findings for the 24-item TSES.

For the fall 12-item TSES data, as shown in Table 5, model fit indices indicated satisfactory fit for the 3-factor model ($\chi^2_{MLR}[51] = 98.75$, $p < .001$, CFI = .97, RMSEA = .06, SRMR = .03), and adequate fit for the 1-factor model, $\chi^2_{MLR}(54) = 168.99$, $p < .001$, CFI = .92, RMSEA = .09, SRMR = .05. As a probable result of smaller sample size, the bifactor model would not converge. Comparative model fit indices informed further
comparison of the models for the fall data, including AIC, BIC, and the chi-square difference test. AIC and BIC values supported the 3-factor model. The chi-square difference test showed the 3-factor model had better fit than then 1-factor model, $\chi^2_{\text{MLR}(3)} = 49.55, p < .001$.

Winter 12-item TSES results were similar to fall data results. Table 5 indicates that the 3-factor model ($\chi^2_{\text{MLR}[51]} = 72.40, p < .001, \text{CFI} = .98, \text{RMSEA} = .05, \text{SRMR} = .04$) indices revealed satisfactory fit; the 1-factor model ($\chi^2_{\text{MLR}[54]} = 123.49, p < .001, \text{CFI} = .93, \text{RMSEA} = .08, \text{SRMR} = .05$) had adequate fit. As with the fall data, the bifactor model did not converge. AIC and BIC indices favored the 3-factor model relative to the 1-factor model, as did the chi-square difference test, $\chi^2_{\text{MLR}(3)} = 83.49, p < .001$.

For the spring 12-item TSES data, the bifactor model converged, likely due to greater sample size. Table 5 indicates that the bifactor model ($\chi^2_{\text{MLR}[42]} = 81.05, p < .001, \text{CFI} = .97, \text{RMSEA} = .06, \text{SRMR} = .03$) had satisfactory fit. The 3-factor model ($\chi^2_{\text{MLR}[51]} = 101.91, p < .001, \text{CFI} = .96, \text{RMSEA} = .06, \text{SRMR} = .04$) had adequate fit; the 1-factor model ($\chi^2_{\text{MLR}[54]} = 162.90, p < .001, \text{CFI} = .92, \text{RMSEA} = .08, \text{SRMR} = .05$) had reasonable fit. The AIC index favored the bifactor model, while the BIC index supported the 3-factor model over the 1-factor model. The chi-square difference test showed the 3-factor model had better fit than the 1-factor model, $\chi^2_{\text{MLR}(3)} = 41.21, p < .001$. Relatively speaking, the 3-factor model had better fit than the 1-factor or bifactor models across the three waves of data for the 12-item TSES, although there were not sufficient numbers to conduct bifactor analyses for two of the data points. As with the 24-item TSES, correlations among factors in the 3-factor model were high and suggested
redundant content for the 12-item TSES. The correlations ranging from .82 to .93 for fall and winter, and from .82 to .95 for spring data.

Table 5

CFA Model Fit Results for 1-Factor, 3-Factor, and Bifactor Models to the 12-item TSES using MLR Estimation

<table>
<thead>
<tr>
<th>Model</th>
<th>χ²_{MLR}</th>
<th>df</th>
<th>RMSEA 90% CI [LL, UL]</th>
<th>CFI</th>
<th>SRMR</th>
<th>AIC</th>
<th>BIC</th>
<th>χ² diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall (n = 265)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-factor</td>
<td>168.99***</td>
<td>54</td>
<td>.09 [.08, .10]</td>
<td>.92</td>
<td>.05</td>
<td>8,294.95</td>
<td>8,423.82</td>
<td></td>
</tr>
<tr>
<td>3-factor</td>
<td>98.75**</td>
<td>51</td>
<td>.06 [.04, .08]</td>
<td>.97</td>
<td>.03</td>
<td>8,201.50</td>
<td>8,341.10</td>
<td>49.55</td>
</tr>
<tr>
<td>Bifactor</td>
<td></td>
<td></td>
<td>Non-convergence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winter (n = 196)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-factor</td>
<td>123.49***</td>
<td>54</td>
<td>.08 [.06, 1.0]</td>
<td>.93</td>
<td>.05</td>
<td>5,968.46</td>
<td>6,086.47</td>
<td></td>
</tr>
<tr>
<td>3-factor</td>
<td>72.40*</td>
<td>51</td>
<td>.05 [.02, .07]</td>
<td>.98</td>
<td>.04</td>
<td>5,901.87</td>
<td>6,029.72</td>
<td>83.49</td>
</tr>
<tr>
<td>Bifactor</td>
<td></td>
<td></td>
<td>Non-convergence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spring (n = 298)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-factor</td>
<td>162.90***</td>
<td>54</td>
<td>.08 [.07, 1.0]</td>
<td>.92</td>
<td>.05</td>
<td>8,865.67</td>
<td>8,998.77</td>
<td></td>
</tr>
<tr>
<td>3-factor</td>
<td>101.91***</td>
<td>51</td>
<td>.06 [.04, .07]</td>
<td>.96</td>
<td>.04</td>
<td>8,776.05</td>
<td>8,920.23</td>
<td>41.21</td>
</tr>
<tr>
<td>Bifactor</td>
<td>81.05**</td>
<td>42</td>
<td>.06 [.04, .07]</td>
<td>.97</td>
<td>.03</td>
<td>8,755.71</td>
<td>8,933.17</td>
<td></td>
</tr>
</tbody>
</table>

Note. TSES = Teachers’ Sense of Efficacy Scale; df = degrees of freedom; RMSEA = root mean square error of approximation; CFI = comparative fit index; SRMR = standardized root mean square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion; χ² diff = chi-square difference.

* p < .05. ** p < .01. *** p < .001.

Factor loadings for the 24- and 12-item TSES. Factor loadings were similar for all waves of data within a given CFA model. For the sake of parsimony, the results for the fall 24- and 12-item TSES are presented in Table 6. Standardized factor loadings for the 1-factor model ranged from .52 to .90; for the 3-factor model, loadings ranged from .50 to .93. For the general factor in the bifactor model, standardized factor loadings ranged from .53 to .83, and for the bifactor specific factors, loadings ranged from -.29 to .48. In addition, analyses included calculation of the average of the differences for all
items between the standardized factor loadings for the 1-factor and bifactor general models, per Reise et al. (2013). The averages of the differences were negligible, ranging from .01 (fall) to .05 (winter) and to .07 (spring), which gave further evidence of a unidimensional model solution being able to provide unbiased parameter estimates if data were multidimensional. As a result, the unidimensional solution was used for all remaining study analyses. To differentiate factor loadings for the 24- and 12-item versions, factor loadings for the 12-item TSES version are bolded in Table 6.
Table 6

1-Factor, 3-Factor, and Bifactor CFA Standardized Factor Loadings for 12- and 24-Item Fall TSES ($n = 265$)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Item Phrase</th>
<th>1-Factor</th>
<th>3-Factor</th>
<th>Bifactor general</th>
<th>Bifactor specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruct</td>
<td>Use varied assessment strategies</td>
<td>.79***</td>
<td>.81***</td>
<td>.77***</td>
<td>-.07</td>
</tr>
<tr>
<td></td>
<td>Provide alternative explanations</td>
<td>.80***</td>
<td>.82***</td>
<td>.77***</td>
<td>.17</td>
</tr>
<tr>
<td></td>
<td>Craft good questions</td>
<td>.73***</td>
<td>.75***</td>
<td>.72***</td>
<td>.35***</td>
</tr>
<tr>
<td></td>
<td>Implement alternative strategies</td>
<td>.84***</td>
<td>.86***</td>
<td>.80***</td>
<td>-.29***</td>
</tr>
<tr>
<td></td>
<td>Respond to difficult questions</td>
<td>.78***</td>
<td>.80***</td>
<td>.74***</td>
<td>.13</td>
</tr>
<tr>
<td></td>
<td>Adjust your lessons for individuals</td>
<td>.81***</td>
<td>.83***</td>
<td>.76***</td>
<td>-.21</td>
</tr>
<tr>
<td></td>
<td>Gauge student comprehension</td>
<td>.85***</td>
<td>.87***</td>
<td>.83***</td>
<td>.22*</td>
</tr>
<tr>
<td></td>
<td>Provide appropriate challenges</td>
<td>.81***</td>
<td>.84***</td>
<td>.80***</td>
<td>-.05</td>
</tr>
<tr>
<td>Manage</td>
<td>Control disruptive behavior</td>
<td>.88***</td>
<td>.90***</td>
<td>.77***</td>
<td>.48***</td>
</tr>
<tr>
<td></td>
<td>Get children to follow class rules</td>
<td>.82***</td>
<td>.84***</td>
<td>.73***</td>
<td>.31**</td>
</tr>
<tr>
<td></td>
<td>Calm disruptive student</td>
<td>.85***</td>
<td>.87***</td>
<td>.75***</td>
<td>.37***</td>
</tr>
<tr>
<td></td>
<td>Establish class management</td>
<td>.85***</td>
<td>.87***</td>
<td>.77***</td>
<td>.26**</td>
</tr>
<tr>
<td></td>
<td>Keep students from ruining lesson</td>
<td>.80***</td>
<td>.82***</td>
<td>.70***</td>
<td>.42***</td>
</tr>
<tr>
<td></td>
<td>Respond to defiant students</td>
<td>.83***</td>
<td>.85***</td>
<td>.75***</td>
<td>.32***</td>
</tr>
<tr>
<td></td>
<td>Clear behavioral expectations</td>
<td>.81***</td>
<td>.83***</td>
<td>.73***</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>Establish routines</td>
<td>.81***</td>
<td>.83***</td>
<td>.74***</td>
<td>.16</td>
</tr>
<tr>
<td>Engage</td>
<td>Help students believe well</td>
<td>.78***</td>
<td>.50***</td>
<td>.70***</td>
<td>.34***</td>
</tr>
<tr>
<td></td>
<td>Help students value learning</td>
<td>.78***</td>
<td>.82***</td>
<td>.74***</td>
<td>.33**</td>
</tr>
<tr>
<td></td>
<td>Motivate low interest students</td>
<td>.80***</td>
<td>.82***</td>
<td>.75***</td>
<td>.39***</td>
</tr>
<tr>
<td></td>
<td>Assist families help children</td>
<td>.59***</td>
<td>.60***</td>
<td>.53***</td>
<td>.29**</td>
</tr>
<tr>
<td></td>
<td>Assist failing student</td>
<td>.78***</td>
<td>.79***</td>
<td>.75***</td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td>Help students think critically</td>
<td>.52***</td>
<td>.54***</td>
<td>.71***</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>Foster student creativity</td>
<td>.70***</td>
<td>.72***</td>
<td>.65***</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>Get through to difficult students</td>
<td>.90***</td>
<td>.93***</td>
<td>.74***</td>
<td>.06</td>
</tr>
</tbody>
</table>

Note. TSES = Teachers’ Sense of Efficacy Scale; CFA = confirmatory factor analysis; **bolded** item phrases and loadings reflect the 12-item TSES.

*p < .05. **p < .01. ***p < .001.

**Reliability analyses.** Internal consistency of reliability of TSES scores was examined with coefficient omega (i.e., a model-based reliability based on CFA model results) for fall, winter, and spring data. Results indicated that, for all data points, the coefficient omega was .97 with a bootstrap corrected ($k = 1,000$) 95% confidence interval
[.96, .97] for the 24-item TSES.

**Longitudinal measurement invariance.** LMI (Jöreskog, 1971) analyses were conducted for all surveys. Altogether, there were 759 TSES surveys from fall, winter, and spring data collection; of that total, 467 surveys had data for the 24 TSE items. However, due to the large amount of variables in the 24-item dataset, several model fit statistics could not be computed (\(\chi^2_{\text{MLR}}, \text{RMSEA, CFI}\)). Therefore, LMI analyses examined the 12-item TSES version, including baseline, metric, and scalar levels of analysis; see results in Table 7. Fit indices (\(\chi^2_{\text{MLR}}, \text{RMSEA, SRMR, and CFI}\)) indicated adequate fit for the baseline model, \(\chi^2_{\text{MLR}} (555) = 1,085.49, p < .001, \text{CFI} = .90, \text{RMSEA} = .05, \text{SRMR} = .08\). Fit indices for the metric invariance model (\(\chi^2_{\text{MLR}} [577] = 1,129.89, p < .001, \text{CFI} = .89, \text{RMSEA} = .05, \text{SRMR} = .08\)) indicated no relative change from the baseline model. One of the model comparison indices (chi-square difference test) indicated a significant difference from the baseline model (\(\Delta \chi^2_{\text{MLR}} = 44.62, p < .001\)), although a difference was not shown for \(\Delta \text{CFI} (.01)\). Since chi-square tests are known to be sensitive to large sample sizes, emphasis was placed on the acceptable range of \(\Delta \text{CFI}\) (Cheung & Rensvold, 2002). Therefore, analyses proceeded to the scalar level, for which the chi-square difference test and \(\Delta \text{CFI}\) showed a loss of fit relative to the metric model, \(\Delta \chi^2_{\text{MLR}}(602) = 1,914.43, p < .001, \Delta \text{CFI} = .35\). As a result, longitudinal analyses assuming scalar invariance should be interpreted with caution.
Table 7

**Longitudinal Measurement Invariance Results for 12-item TSES scores (n = 467) using MLR Estimation**

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2_{MLR}$</th>
<th>$\Delta \chi^2$</th>
<th>df</th>
<th>CFI</th>
<th>$\Delta$CFI</th>
<th>SRMR</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>1,085.49</td>
<td></td>
<td>555</td>
<td>.90</td>
<td>.08</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Metric</td>
<td>1,129.89</td>
<td>44.62</td>
<td>577</td>
<td>.89</td>
<td>.01</td>
<td>.08</td>
<td>.05</td>
</tr>
<tr>
<td>Scalar</td>
<td>3,002.21</td>
<td>1,914.43</td>
<td>602</td>
<td>.54</td>
<td>.35</td>
<td>2.05</td>
<td>.09</td>
</tr>
</tbody>
</table>

*Note.* TSES = Teachers’ Sense of Efficacy Scale; $\Delta \chi^2_{MLR}$ = change in $\chi^2$; df = degrees of freedom; CFI = comparative fit index; $\Delta$CFI = change in CFI; RMSEA = root mean square error of approximation. All models were statistically significant at $p < .001$.

**Results for Statistical Analyses**

**Longitudinal growth modeling for the TSES.** Longitudinal growth modeling (McArdle, 1988; Meredith & Tisak, 1990) of TSES scores was conducted in *Mplus* 7.4 (Muthén & Muthén, 1998-2015) with common residual variance estimated. After listwise deletion of missing data from the 467 surveys with TSE data, the sample size was 439 for growth modeling analyses.

First, analyses were completed for growth (i.e., item sum scores) models for both 12- and 24-item TSES versions. Table 8 summarizes the growth curve modeling results using the sum scores as the dependent for the 12-item TSES variable; Table 9 presents results for growth curve modeling for the 24-item TSES variable. Model fit indices indicated strong fit for Model 1 (i.e., TSES growth scores only). For both 12- and 24-item versions, comparisons between each subsequent model did not show improvement in fit from Model 1 to Model 2 (i.e., TSES growth and teacher demographics at intercept) or from Model 2 to Model 3 (i.e., TSES growth, teacher demographics at intercept, interaction of time with teacher demographic variables); see Tables 8 and 9. The results
show that fit worsened when teacher demographics were added to the model at the intercept level.

Table 8

*Growth Model Fit Statistics Using Sum Scores for the 12-Item TSES based on MLR Estimation (n = 439)*

<table>
<thead>
<tr>
<th>Model</th>
<th>χ²_{MLR}</th>
<th>df</th>
<th>Δdf</th>
<th>RMSEA</th>
<th>CFI</th>
<th>SRMR</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-Growth</td>
<td>0.62</td>
<td>3</td>
<td></td>
<td>&lt; .001</td>
<td>1.0</td>
<td>.07</td>
<td>1,722.60</td>
<td>1,747.11</td>
</tr>
<tr>
<td>2-Growth, demo</td>
<td>25.79</td>
<td>17</td>
<td>14</td>
<td>.03</td>
<td>.93</td>
<td>.04</td>
<td>1,719.22</td>
<td>1,772.32</td>
</tr>
<tr>
<td>3-Growth, demo &amp; inter</td>
<td>7.69</td>
<td>0</td>
<td>7</td>
<td>&lt; .001</td>
<td>1.0</td>
<td>.03</td>
<td>1,715.23</td>
<td>1,733.45</td>
</tr>
</tbody>
</table>

*Note.* TSES = Teachers’ Sense of Efficacy Scale; 1-Growth = sum score growth; 2-Growth, demo = sum score growth, teacher demographics; 3-Growth, demo & inter = sum score growth, teacher demographics, interactions; df = degrees of freedom; Δdf = degrees of freedom change; RMSEA = root mean square error of approximation; CFI = comparative fit index; SRMR = standard root mean square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion.

Table 9

*Growth Model Fit Statistics Using Sum Scores for the 24-Item TSES based on MLR Estimation (n = 439)*

<table>
<thead>
<tr>
<th>Model</th>
<th>χ²_{MLR}</th>
<th>df</th>
<th>Δdf</th>
<th>RMSEA</th>
<th>CFI</th>
<th>SRMR</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-Growth</td>
<td>1.30</td>
<td>3</td>
<td></td>
<td>&lt; .001</td>
<td>1.0</td>
<td>.09</td>
<td>1,674.15</td>
<td>1,698.65</td>
</tr>
<tr>
<td>2-Growth, demo</td>
<td>23.71</td>
<td>17</td>
<td>14</td>
<td>&lt; .001</td>
<td>.95</td>
<td>.05</td>
<td>1,668.27</td>
<td>1,721.37</td>
</tr>
<tr>
<td>3-Growth, demo &amp; inter</td>
<td>8.12</td>
<td>10</td>
<td>7</td>
<td>&lt; .001</td>
<td>1.0</td>
<td>.05</td>
<td>1,667.01</td>
<td>1,748.70</td>
</tr>
</tbody>
</table>

*Note.* TSES = Teachers’ Sense of Efficacy Scale; 1-Growth = sum score growth; 2-Growth, demo = sum score growth, teacher demographics; 3-Growth, demo & inter = sum score growth, teacher demographics, interactions; df = degrees of freedom; Δdf = degrees of freedom change; RMSEA = root mean square error of approximation; CFI = comparative fit index; SRMR = standard root mean square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion.

The change in TSES scores across the preschool year also was investigated by examining the Model 1 correlation between slope and intercept, and the mean slope. The
change across the year was slight. The correlation between the slope and intercept was -.49, \( p < .001 \) and the mean slope was .41, \( p < .05 \), indicating that teachers who started with higher TSES scores increased their scores at a slower rate, teachers who started with lower TSES scores increased more quickly, and overall, teacher TSES growth was small.

Next, analyses were completed using a latent growth modeling approach (i.e., multiple indicators growth curve model) for the 12-item TSES. Table 10 summarizes the global model fit statistics for each model for the 12-item TSES. As with growth modeling based on sum scores, model fit indices indicated strong fit for Model 1, with no improvement in fit for Models 2 or 3 for the latent growth models. Actually, fit worsened when teacher demographics were added to the latent growth model. For the 12-item latent growth model with no teacher demographic variables added (Model 1 in Table 10), the correlation between slope and intercept was -.47, \( p < .01 \) and the mean slope was .30, \( p < .05 \), thus mirroring the results from the growth model with slightly more conservative estimates. For the 24-item latent growth model, convergence did not occur.
Table 10

*Latent or Multiple Indicator Growth Model Fit Results for the 12-item TSES (n = 439)*

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2_{\text{MLR}}$</th>
<th>df</th>
<th>$\Delta$df</th>
<th>RMSEA</th>
<th>CFI</th>
<th>SRMR</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-Growth</td>
<td>1,343.64</td>
<td>635</td>
<td>.05</td>
<td>.86</td>
<td>.09</td>
<td></td>
<td>22,059.95</td>
<td>22,333.61</td>
</tr>
<tr>
<td>2-Growth demo</td>
<td>1,632.85</td>
<td>880</td>
<td>245</td>
<td>.04</td>
<td>.86</td>
<td>.09</td>
<td>22,056.39</td>
<td>22,358.64</td>
</tr>
<tr>
<td>3-Growth demo &amp; inter</td>
<td>1,612.55</td>
<td>873</td>
<td>7</td>
<td>.04</td>
<td>.86</td>
<td>.08</td>
<td>22,049.88</td>
<td>22,380.72</td>
</tr>
</tbody>
</table>

*Note.* TSES = Teachers’ Sense of Efficacy Scale; 1-Growth = factor score growth; 2-Growth demo = factor score growth, teacher demographics; 3-Growth demo & inter = factor score growth, teacher demographics, interactions; $df$ = degrees of freedom; $\Delta df$ = change in degrees of freedom; RMSEA = root mean square error of approximation; CFI = comparative fit index; SRMR = standard root mean square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion. All models were statistically significant at $p < .001$.

These results indicated that the growth model without predictors fit the data, and that adding predictors to the model for intercept only (Model 2) and for time (Model 3) did not improve model fit. As such, it can be concluded that none of the predictors helped explain growth. Because of findings from CFA, longitudinal measurement invariance testing, and growth modeling, scores from the first TSES administration (i.e., fall) were used as the measure of TSE in all subsequent analyses.

**HLM Results for Child Literacy Outcomes.** HLM analyses for child literacy scores were conducted in HLM 7.01. Descriptive results for the three HLM files are presented in Table 11. Level 1 included all children with literacy growth scores ($n = 12,304$); Level 2 included all children with complete background data ($n = 9,674$). Level 3 included fall teachers with complete demographic and TSES data ($n = 240$); for the final analyses, there were 204 cases with complete teacher data that also had child literacy growth data.
HLM analyses examined fixed effects, random effects, proportion of variance, effect size, and correlations between intercept and growth parameters. HLM analyses indicated that child variables of gender, race and ethnicity, special education status, and age were significant for entry level literacy scores. Children who were girls, White, not enrolled in special education, and older preschoolers had higher literacy scores (by 2.6, 1.01, 1.01, and 1.77 points, respectively, holding other variables constant) in fall 2013; see Table 11. The average literacy entry score for children who were boys, White, not in special education, and older preschoolers was 28.34 out of a possible 108 points. When examining child literacy growth over time, all child variables were significant, following the patterns indicated at entry. Children who were girls, White, not in special education, and older preschoolers increased literacy scores at a greater rate (.18, .14, .08, and .70 points per month in preschool) than did children who were boys, non-White, in special education, and younger preschoolers. The average growth in literacy score was 3.64 points for each unit increase in time; thus, for each month in preschool, the average literacy growth was 3.64 points, while holding all other variables constant.

Two teacher variables related significantly to child literacy growth: IECE certification and TSES scores. Teachers with IECE certification and with higher TSES scores had children with more growth in literacy scores, than did teachers without an IECE and with lower TSES scores. Effect sizes were trivial (.01) for fall TSES scores at entry; for child literacy growth, effect sizes were small (.26) for teacher certification and trivial (.004) for fall TSES scores. The effect size for teacher certification for child literacy growth may be described as follows: for each unit of change in the TSES score, it was associated with .26 standard deviation of the child literacy growth. Effect sizes for
TSES, at entry and for growth, were so small as to be trivial. See Table 11 for detailed results of the HLM findings. Pearson’s correlations between literacy intercept and growth parameters were .99 at entry, and .97 for the final model.

Table 11

Estimates for HLM Analyses of Child Literacy Growth in Preschool

<table>
<thead>
<tr>
<th></th>
<th>Child Literacy Growth</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Effect</td>
<td>SE</td>
</tr>
<tr>
<td>Average within-child literacy scores</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At entry</td>
<td>28.34***</td>
<td>.60</td>
</tr>
<tr>
<td>For growth</td>
<td>3.64***</td>
<td>.10</td>
</tr>
<tr>
<td>Effects of between-child variables at entry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>2.60***</td>
<td>.25</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td>-1.01*</td>
<td>.40</td>
</tr>
<tr>
<td>IEP status</td>
<td>-1.77**</td>
<td>.49</td>
</tr>
<tr>
<td>Age</td>
<td>1.01**</td>
<td>.03</td>
</tr>
<tr>
<td>Effects of between-child variables for rate of growth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>0.18***</td>
<td>.04</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td>-0.14*</td>
<td>.06</td>
</tr>
<tr>
<td>IEP status</td>
<td>-0.70***</td>
<td>.06</td>
</tr>
<tr>
<td>Age</td>
<td>0.08***</td>
<td>.01</td>
</tr>
<tr>
<td>Effects of between-teacher variables at entry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall TSES</td>
<td>0.06*</td>
<td>.02</td>
</tr>
<tr>
<td>Effects of between-teacher variables for rate of growth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IECE</td>
<td>.54**</td>
<td>.19</td>
</tr>
<tr>
<td>Fall TSES</td>
<td>.01*</td>
<td>.00</td>
</tr>
</tbody>
</table>

Note. SE = standard error; d = effect size; IEP = Individual Education Plan; TSES = Teachers’ Sense of Efficacy Scale; IECE = Interdisciplinary Early Childhood Education.

*p < .05. **p < .01. ***p < .001.

The intraclass correlation coefficients (ICC) indicated the degree of unexplained variance for child and teacher variables in the baseline Model 1 (i.e., no predictors added). At entry (i.e., initial status for child literacy), ICC values indicated that modeled child variables accounted for 64% and modeled teacher variables accounted for 36% of unexplained variance in child literacy scores. For child literacy growth, child variables accounted for 35% and teacher variables accounted for 65% of unexplained variance in literacy growth.
HLM analyses indicated the proportions of explained variance for the baseline and final models (Raudenbush & Bryk, 1988) for child literacy scores. The baseline Model 1 included child literacy growth with no predictors added; Model 2 added child background variables; Model 3 added teacher characteristics; and Model 4 was the final HLM model with significant variables. Modeled child variables accounted for the largest proportion of explained variance, including 43% at entry and 60% of the variance for literacy growth scores. Modeled teacher variables accounted for a small proportion of the explained variance in literacy scores, with 2% at entry and 4% of growth in child literacy scores. Thus, modeled child variables accounted for the greater proportion of explained variance while modeled teacher variables accounted for a small proportion of explained variance for child literacy initial status and growth in literacy scores. See Table 12 for findings.

Table 12

HLM Unexplained and Explained Variance

<table>
<thead>
<tr>
<th>Model</th>
<th>Proportion of Unexplained Variance</th>
<th>Proportion of Explained Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intercept</td>
<td>Slope</td>
</tr>
<tr>
<td></td>
<td>Child</td>
<td>Teacher</td>
</tr>
<tr>
<td>Baseline Model 1</td>
<td>108.30</td>
<td>60.77</td>
</tr>
<tr>
<td>Child variables Model 2</td>
<td>61.54</td>
<td>62.41</td>
</tr>
<tr>
<td>Teacher variables Model 3</td>
<td>109.81</td>
<td>59.45</td>
</tr>
<tr>
<td>Final Model 4</td>
<td>61.53</td>
<td>60.68</td>
</tr>
<tr>
<td>Model 2 (child)</td>
<td>.43</td>
<td>.60</td>
</tr>
<tr>
<td>Model 3 (teacher)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Baseline Model 1 = child literacy growth only; Model 2 = child literacy and background variables; Model 3 = child literacy and background; teacher characteristics; Model 4 = significant variables only.
Chapter 5: Conclusions and Limitations

The current study conducted extensive psychometric analyses of a widely used measure of teacher self-efficacy, the TSES, and conducted HLM analyses of teacher characteristics in relation to preschool child literacy growth. The goals of the study included assessing the psychometric quality of the TSES, and investigating relationships between preschool teacher characteristics and child literacy growth. With regard to the first research question, CFA and ancillary bifactor analyses indicated that a unidimensional structure was appropriate and would not result in biased parameter estimates. The second research question pertained to longitudinal measurement invariance and longitudinal growth of TSES scores. Analyses indicated evidence for metric measurement invariance for the 12-item TSES, and growth in TSES scores across the academic year was slight. The third research question addressed the relationship between preschool teacher characteristics and child literacy growth; results indicated that teacher certification and TSES scores were significant to child literacy growth.

Conclusions

Findings from the current study indicate six areas for consideration in further research. First, additional psychometric examination of the TSES is indicated (Pajares, 1992; Tschannen-Moran & Woolfolk Hoy, 2001), including factor structure, item difficulty and discrimination, and LMI analyses. Second, development of measures that more accurately reflect preschool teacher self-efficacy for child literacy is indicated, as the TSES is not specific to preschool teachers or child literacy growth. Third, study findings indicate that TSES scores did not change much over one academic year; further studies can investigate the range of TSES score growth in samples over time. Fourth, investigation of ways to enhance TSE is important for teacher pre- and in-service training
programs. Fifth, findings indicate the need for emphasizing the positive relationship found between early childhood teacher certification and child literacy growth, as documented for the first time in this study. Sixth, in light of national attention on improved child literacy, identification of additional teacher characteristics that relate to child literacy growth, especially for lower achieving children, is encouraged, to clarify the teacher characteristics that support this important area of child development.

The first conclusion, based on study findings, is that additional psychometric analyses of the TSES are indicated. While the current study supported a unidimensional model for TSES scores, as did Khairani & Razak (2012), other studies found evidence for unidimensional and 3-factor models (Brown, 2005; Tschannen-Moran & Woolfolk Hoy, 2001). Investigation of dimensionality with other samples is needed. In addition, further use of psychometric test methods such as item response theory (IRT; Embretson & Reise, 2000) is recommended to add to the knowledge base for item and response category properties of the TSES. Psychometric analyses of measures including IRT modeling result in greater predictive accuracy and reliability of instruments such as the TSES (Borsboom, 2006; Kim & Cimilli, 2014; Millsap, 2010). As Kim and Cimilli (2014) indicated, few studies have examined preschool child literacy or the measurement properties of assessment instruments; IRT modeling of longitudinal growth can result in accurate parameter estimates of child literacy skills. Preliminary IRT analyses of the TSES indicated that the unidimensional model provided the most complete and precise information for TSES (Gooden, Li, Toland, & Danner, 2015). Further such study with the TSES will add to the knowledge base for measurement precision of the instrument.

Recommended psychometric analyses for future studies also include additional LMI analyses of TSES items. Ignoring non-invariance of measures can decrease the
accuracy and validity of research findings (Chen & West, 2008; Millsap, 2010). Studies have demonstrated that item non-invariance can lead to inaccurate means for latent variables and inaccurate estimations of interaction effects. While current study results indicated tenable support for metric invariance of the 12-item TSES, weak partial invariance (i.e., removing items with the greatest non-invariance) testing may result in a version with improved invariance. Investigations with smaller amounts of missing data and variables may permit LMI analyses to converge with the 24-item TSES. Further, additional LMI modeling with subsets of the sample (i.e., teachers with varying educational levels) will indicate the extent of invariance across specific groups of teachers.

A second conclusion, based on study findings, is that the development of an instrument that measures specific preschool TSE for child literacy may result in the identification of a larger association between TSE and child literacy outcomes. The TSES measures general instruction, classroom management, and child engagement for teachers from elementary through high school levels, and has been used with preschool populations (Brown, 2005). Measurement of TSE for child literacy may need indicators that are specific to preschool literacy, such as phonemic, alphabet, and print knowledge. Some examples of content-specific measures of self-efficacy include the Mathematics Self-Efficacy Scale (Betz & Hackett, 1983) and the Longitudinal Assessment of Engineering Self-Efficacy (Marra & Bogue, 2006). Development of a specific TSE measure with a closer conceptual match between preschool TSE and child literacy may reveal a larger association between these variables.

Third, study results indicate the need for further investigation of the stability of TSES scores, as no other preschool studies examined changes in TSES scores over time.
While the current study found a limited range and little change of TSES scores over one academic year, disaggregated analyses by subsets of the sample may reveal more information about the variability in TSES. In addition, it is advisable to investigate whether possible ceiling effects of TSES items limited growth in scores.

A fourth conclusion that is important to the field of educational psychology is the need for further investigation of factors that enhance TSE. Examination of factors and interventions that enhance teachers’ efficacy to teach literacy is important to improved outcomes for young children. Based on review of literature, focused teacher training in the areas of mastery experiences and vicarious learning are suggested (Bandura, 1993; Pajares, 1992; Usher & Pajares, 2008) for improved literacy instruction. Systematic initiatives at the pre-and in-service levels of teacher training can ensure mastery and vicarious learning experiences for new and returning teachers. Measurement of TSE prior to and after intervention programs designed to improve teachers’ abilities to increase child literacy outcomes may reveal specific factors that enhance teachers’ TSE. Barnett (2003), Early et al. (2007), and Pianta et al. (2005) emphasized the need for targeted pre-service teacher education and on-the-job coaching to improve classroom instruction and child outcomes.

A fifth conclusion, based on study findings, is the need to highlight the relationship found between early childhood teacher certification and preschool child literacy outcomes. This study found that teacher certification had a small, positive association with child literacy outcomes. Prior preschool studies had not specifically examined associations between teacher certification and child literacy outcomes (Early et al., 2007; Pianta et al., 2005). The implications of current study findings for teacher certification programs are important, as are early childhood program requirements for
teacher certification. Findings suggest that increasing the requirement for IECE certification across early childhood programs would be associated with increased child literacy scores throughout the state. Incentives such as scholarships, on-the-job training, and mentoring improve teacher certification rates (Guskey, 1986; Pianta et al., 2005; Trivette et al., 2012; Usher & Pajares, 2008). In Kentucky, incentives are in place through the KIDS (Kentucky Invests in Developing Success) NOW Initiative (Kentucky Governor's Early Childhood Task Force, 1999) for state-funded preschool teachers to complete university degrees and certifications; the current study provides support for extending this program to all early childhood agencies.

The sixth and final conclusion, based on study results, is that there is a need to identify other teacher variables that contribute to the variance in child literacy growth and that address instruction for lower achieving children. Most of the modeled teacher characteristics included in the TSES did not relate to measured child literacy outcomes. Further, as indicated in Table 11, children who were non-White and who received special education services had lower initial literacy scores and grew at slower rates. Given the national emphasis on improving child literacy and reducing achievement gaps (NCLB Act of 2002; Pianta, 2003; Shonkoff & Phillips, 2000), identifying other teacher characteristics that improve child literacy outcomes, especially for children who are achieving at slower rates, is imperative. Research efforts and teacher training programs can identify and support teacher characteristics that facilitate literacy instruction, with special attention to strategies for closing achievement gaps.

Limitations and Future Directions

Results indicate four study limitations that indicate directions for future research.
Limitations include the relatively small and homogenous sample size, limited demographic information about Kentucky’s preschool teacher population, missing data, and the use of one self-report measure of TSE. Sample size was limited, with response rates of 39%, 29%, and 44%. In addition, representativeness of the sample could not be determined. By comparing study demographics with state and national census data, it was apparent that the current sample had greater numbers of children with special education services, and disproportionate numbers of children from some racial groups. The sample had fewer White and more mixed and other races than in Kentucky, and had fewer White and more African American children than nationally. The larger proportion of children with special education services is not surprising, given the eligibility criteria for state-funded programs that included any child aged three or four years with an Individual Education Plan (Kentucky Governor's Early Childhood Task Force, 1999). Future studies with nationally representative and more diverse preschool samples would increase measurement precision and the potential for generalizability.

Another limitation included missing data at the child and teacher levels, which limited analyses that would converge. This limitation reflects the study design that used preschool assessment data designed to inform instruction that was already collected by teachers. As a result, there was considerable missing data across waves of data collection and for individual teachers. CFAs would not converge for the bifactor model for two of the data points due to smaller sample sizes, and LMI analyses would not converge for the 24-item TSES. Larger sample sizes and less missing data would allow for convergence of both types of analyses, and more complete psychometric information about TSES scores.
Lastly, the scope of the study did not permit concurrent examination with another measure of self-efficacy, such as Bandura’s TES (1997) scale, or with a measure not biased by teacher self-report. The current study selected one instrument to reduce teacher burden and to increase response rates, which limited comparative analyses with other measures of TSE. Prior studies compared results for more than one TSE measure, allowing for correlations between scales and factors (Ashton et al., 1982; Guskey, 1981). Future studies that include administration of the TSES and other TSE instruments would allow for expanded study of the dimensionality and measurement of TSE.
Appendices

Appendix A: IRB Approval
Appendix B: TSES

The purpose of this online survey is to learn more about the kinds of things that are challenging to preschool teachers. This survey was developed by Tschannen-Moran and Woolfolk Hoy in 2001, to help gain a better understanding of the kinds of things that create challenges for teachers.

1. What is the location of your classroom? (Select one)
   1. Elementary school
   2. Early childhood center
   3. Blended Head Start/state-funded preschool
   4. Head Start center

2. With which of the following do you identify? (Select one)
   1. African American
   2. Asian American
   3. Hispanic American
   4. Other
   5. White, Non-Hispanic American

3. What is the zip code of your school? (Enter 5-digit number)

4. What is the approximate proportion of students who receive free or reduced lunch in your class? (Select one)
   1. 1-20%
   2. 21-40%
   3. 41-60%
   4. 61-80%
   5. 81-100%

5. How many years have you taught preschool? (Enter number; for example, 1, 2, etc.)

6. How much professional development have you received since July 2013? (Select one)
   1. Less than 1 hour
   2. 1 hour
   3. 2-5 hours
   4. 6-10 hours
   5. More than 10 hours

7. Check if you have any of the following teaching certificates (Check all that apply).
   1. Interdisciplinary Early Childhood Education (IECE)
   2. Elementary Education
   3. Child Development Associate (CDA)
   4. Special Education
8. What is the highest level of education you have achieved? (Select one)
1. High school diploma or GED
2. Associate Degree (AA)
3. Bachelor’s Degree (BA or BS)
4. Master’s Degree (MA or MS)

Please answer the next 24 questions, using the 9-point scale indicated, by considering the combination of your ability, resources, and opportunity to do each of the following in your present position. Select one answer option for each of the 24 questions.

<table>
<thead>
<tr>
<th>Item</th>
<th>Response Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How much can you do to get through to the most difficult students?</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>2. How much can you do to help your students think critically?</td>
<td></td>
</tr>
<tr>
<td>3. How much can you do to control disruptive behavior in the classroom?</td>
<td></td>
</tr>
<tr>
<td>4. How much can you do to motivate students who show low interest in school work?</td>
<td></td>
</tr>
<tr>
<td>5. To what extent can you make your expectations clear about student behavior?</td>
<td></td>
</tr>
<tr>
<td>6. How much can you do to get students to believe they can do well in school?</td>
<td></td>
</tr>
<tr>
<td>7. How well can you respond to difficult questions from your students?</td>
<td></td>
</tr>
<tr>
<td>8. How well can you establish routines to keep activities running smoothly?</td>
<td></td>
</tr>
<tr>
<td>9. How much can you do to help your students value learning?</td>
<td></td>
</tr>
<tr>
<td>10. How much can you gauge student comprehension of what you have taught?</td>
<td></td>
</tr>
</tbody>
</table>
11. To what extent can you craft good questions for your students?
12. How much can you do to foster student creativity?
13. How much can you do to get children to follow classroom rules?
14. How much can you do to improve the understanding of a student who is failing?
15. How much can you do to calm a student who is disruptive or noisy?
16. How well can you establish a classroom management system with each group of students?
17. How much can you do to adjust your lessons to the proper level for individual students?
18. How much can you use a variety of assessment strategies?
19. How well can you keep a few problem students from ruining an entire lesson?
20. To what extent can you provide alternative explanation or example when students are confused?
21. How well can you respond to defiant students?
22. How much can you assist families in helping their children do well in school?
23. How well can you implement alternative strategies in your classroom?
24. How well can you provide appropriate challenges for very capable students?
Appendix C: Teaching Strategies GOLD

Objectives for Development & Learning

Social-Emotional
1. Regulates own emotions and behaviors
   a. Manages feelings
   b. Follows limits and expectations
   c. Takes care of own needs appropriately
2. Establishes and sustains positive relationships
   a. Forms relationships with adults
   b. Responds to emotional cues
   c. Interacts with peers
   d. Makes friends
3. Participates cooperatively and constructively in group situations
   a. Balances needs and rights of self and others
   b. Solves social problems

Physical
4. Demonstrates traveling skills
5. Demonstrates balancing skills
6. Demonstrates gross-motor manipulative skills
7. Demonstrates fine-motor strength and coordination
   a. Uses fingers and hands
   b. Uses writing and drawing tools

Language
8. Listens to and understands increasingly complex language
   a. Comprehends language
   b. Follows directions
9. Uses language to express thoughts and needs
   a. Uses expanding expressive vocabulary
   b. Speaks clearly
   c. Uses conventional grammar
   d. Tells about another time or place
10. Uses appropriate conversational and other communication skills
    a. Engages in conversations
    b. Uses social rules of language

Cognitive
11. Demonstrates positive approaches to learning
    a. Attends and engages
    b. Persists
    c. Solves problems
    d. Shows curiosity and motivation
    e. Shows flexibility and inventiveness in thinking
12. Remembers and connects experiences
    a. Recognizes and recalls
    b. Makes connections
13. Uses classification skills
14. Uses symbols and images to represent something not present
    a. Think symbolically
    b. Engages in sociodramatic play

Literacy
15. Demonstrates phonological awareness
    a. Notices and discriminates rhyme
    b. Notices and discriminates alliteration
    c. Notices and discriminates smaller and smaller units of sound
16. Demonstrates knowledge of the alphabet
    a. Identifies and names letters
    b. Uses letter-sound knowledge
17. Demonstrates knowledge of print and its use
    a. Uses and appreciates books
    b. Uses print concepts
18. Comprehends and responds to books and other text
    a. Interacts during read-alouds and book conversations
    b. Uses emergent reading skills
    c. Retells stories
19. Demonstrates emergent writing skills
    a. Writes names
    b. Writes to convey meaning
Objectives for Development & Learning, continued

Mathematics
20. Uses number concepts and operations
   a. Counts
   b. Quantifies
   c. Connects numerals with their quantities
21. Explores and describes spatial relationships and shapes
   a. Understands spatial relationships
   b. Understands shapes
22. Compares and measures
23. Demonstrates knowledge of patterns

Science and Technology
24. Uses scientific inquiry skills
25. Demonstrates knowledge of the characteristics of living things
26. Demonstrates knowledge of the physical properties of objects and materials
27. Demonstrates knowledge of Earth's environment
28. Uses tools and other technology to perform tasks

Social Studies
29. Demonstrates knowledge about self
30. Shows basic understanding of people and how they live
31. Explores change related to familiar people or places
32. Demonstrates simple geographic knowledge

The Arts
33. Explores the visual arts
34. Explores musical concepts and expression
35. Explores dance and movement concepts
36. Explores drama through actions and language

English Language Acquisition
37. Demonstrates progress in listening to and understanding English
38. Demonstrates progress in speaking English
Appendix D: TSES Invitation Email

Greetings Preschool Coordinators/Special Education Directors,

With the endorsement of the state department of education, I am asking your assistance in distributing this email to all preschool teachers in your district. This survey is being distributed to all districts that use GOLD® as their primary assessment. Please forward this email, with the link included to an electronic survey, to all of your preschool teachers.

Thank you for your assistance in encouraging the completion of this survey! It will help inform preschool teacher practice.

The survey must be completed within 2 weeks of today’s date. Thank you!

Dear Preschool Teachers,

Thank you for educating our youngest citizens!

With the endorsement of the state department of education, preschool teachers across the state are being invited to participate in this survey. Upon completion of this survey in fall, winter, and spring of this year, you will be entered in a random drawing for one of four $75 gift cards. The survey will take approximately 20 minutes. This survey is part of doctoral work for Caroline Gooden, Educational Psychology student at UK; Dr. Fred Danner is her advisor.

If you have any questions, please Caroline Gooden at caroline.gooden@uky.edu or at 859-257-2081.

Please click on the following link: https://uky.az1.qualtrics.com/ to proceed to the survey; the survey must be completed within 2 weeks of today’s date. Thanks for your time and ideas!
References


children’s academic skills: Results from seven studies of preschool programs.

*Child Development, 78,* 558–580. doi:org/10.1111/j.1467-8624.2007.01014.x


doi:10.1002/9781118856406


Gooden, C., Li, Z., Toland, M. D., & Danner, F. (2015). *Item response theory (IRT) examination of Teachers’ Sense of Efficacy Scale (TSES).* Poster session
presented at the meeting of the American Psychological Association Annual Conference, Toronto, Canada.


doi:10.1191/026565901680666527


doi:org/10.4135/9781412983778


NCEDL (2001). *Identifying letters*. Frank Porter Graham Child Development Institute, University of North Carolina at Chapel Hill.

*Findings for children up to age 4 1/2 years (05-4318).* Washington, DC: U.S.

Government Printing Office.


Pintrich (Eds.), *Advances in motivation and achievement, 10* (pp. 1-49).

Greenwich, CT: JAI Press.


doi:org/10.3102/00346543062003307

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C.,


Pianta, R. C. (1992). *The Student-Teacher Relationship Scale.* Charlottesville, VA:

University of Virginia.


Quinn, H. (2014). *Bifactor models, explained common variance (ECV), and the usefulness of scores from unidimensional item response theory analyses* (Master’s thesis). Retrieved from https://cdr.lib.unc.edu/record/uuid:efa83eb6-add5-4d21-a252-823ca4a043f3


Vita

Caroline J. Gooden

Education

1985  Rank 1 Administration of Special Education, Eastern KY University, Richmond, Kentucky
1985  Elementary/Learning and Behavior Disorders Teaching Certification, University of Kentucky
1979  Masters of Science, Early Childhood Special Education, University of Kentucky
1976  Bachelor of Arts, Psychology Major, Drew University, Madison, New Jersey

Professional Experiences

2005-Present  Human Development Institute Research and Development Associate; Co-Principle Investigator: KY Department of Education, First Steps Projects; IDEA Data Center Part C Exiting Co-Lead, University of Kentucky, Lexington, Kentucky.
2005-Present  KY Early Care and Education Level 5 Certified Trainer, Lexington, Kentucky.
2004-2005  People Understanding Special Handicaps Early Childhood Development Center, Executive Director, Frankfort, Kentucky.
2000-2004  Eastern Kentucky University and Lexington Community College Adjunct Faculty, Richmond and Lexington, Kentucky.
1988-2004  Child Development Centers of the Bluegrass Madison County Director, Developmental Interventionist, Head Teacher, Primary Evaluator, Richmond and Lexington, Kentucky.

Professional Publications


Scholastic and Professional Honors


